J Autism Dev Disord (2017) 47:243–259
DOI 10.1007/s10803-016-2949-4

ORIGINAL PAPER

An Evaluation of a Behaviorally Based Social Skills Group for Individuals Diagnosed with Autism Spectrum Disorder

Justin B. Leaf[1] · Jeremy A. Leaf[1] · Christine Milne[1] · Mitchell Taubman[1] · Misty Oppenheim-Leaf[1] · Norma Torres[1] · Donna Townley-Cochran[1] · Ronald Leaf[1] · John McEachin[1] · Paul Yoder[2] · Autism Partnership Foundation

Published online: 2 November 2016

Abstract In this study we evaluated a social skills group which employed a progressive applied behavior analysis model for individuals diagnosed with autism spectrum disorder. A randomized control trial was utilized; eight participants were randomly assigned to a treatment group and seven participants were randomly assigned to a waitlist control group. The social skills group consisted of 32, 2 h sessions. Teachers implemented a variety of behaviorally based procedures. A blind evaluator measured participants' behavior immediately prior to intervention, immediately following intervention, and during 16 and 32-week maintenance probes. Results of the study demonstrated that participants made significant improvements with their social behavior ($p < .001$) following intervention, and the results were maintained up to 32 weeks after intervention had concluded.

Keywords Autism · Applied behavior analysis · Progressive ABA · Social skills · Social skills groups

Introduction

Social skills groups (SSGs) are an intervention strategy in which three or more students, sometimes including students diagnosed with autism spectrum disorder (ASD), come together and are simultaneously taught a variety of social behaviors. SSGs have been found to be effective in teaching a wide variety of behaviors, including: social interaction (e.g., Kamps et al. 1992), greetings (e.g., Barry et al. 2003), handling disagreements (e.g., Laugeson et al. 2012), sportsmanship (e.g., Laugeson et al. 2009), and changing the game when bored (e.g., Kassardjian et al. 2014). There are many potential benefits for implementing SSGs for individuals diagnosed with ASD; these benefits include possible increased observational learning (e.g., Leaf et al. 2013), placing peers in closer proximity to each other, possible promotion of generalization (e.g., Sartini et al. 2013), closer resemblance to typical classroom settings, and more efficient instruction for teachers working with individuals diagnosed with ASD (e.g., Leaf et al. 2013).

This manuscript was presented at the Meeting of the Association of Behavior Analysis in San Antonio Texas in 2015.

✉ Justin B. Leaf
Jblautpar@aol.com

Jeremy A. Leaf
Jeremylautpar@aol.com

Christine Milne
Cmautpar@aol.com

Mitchell Taubman
Mtautpar@aol.com

Misty Oppenheim-Leaf
molbtlc@aol.com

Norma Torres
NTorresautpar@aol.com

Donna Townley-Cochran
dcochranautpar@aol.com

Ronald Leaf
Rlautpar@aol.com

John McEachin
Jmautpar@aol.com

Paul Yoder
paul.yoder@vanderbilt.edu

[1] Autism Partnership Foundation, 200 Marina Drive, Seal Beach, CA 90740, USA

[2] Vanderbilt Peabody College, 416B One Magnolia, Circle VU Mailbox, 228, Nashville, TN, USA

The potential advantages may have led to the widespread clinical implementation of SSGs; today, it is one of the five most commonly implemented interventions to improve social behavior for individuals diagnosed with ASD (IAN Research Findings 2011). In addition to the increase of clinical implementation of SSGs there has been an increase in the number of empirical investigations evaluating SSGs for individuals diagnosed with ASD (Matson et al. 2007). These empirical investigations have included descriptive analyses (e.g., Sartini et al. 2013), single subject designs (e.g., Kamps et al. 1992; Kassardjian et al. 2014), and group designs (e.g., Laugeson et al. 2009, 2012).

Studies that used single-subject designs include Kamps et al. (1992), Kassardjian et al. (2014) and Kohler et al. (2001). In their seminal study, Kamps and colleagues (1992) evaluated the effects of a SSG, implemented in a first grade classroom, to increase social interactions for three individuals diagnosed with ASD. In this study, the authors used procedures based upon the principles of behavior analysis to teach a variety of social behaviors (e.g., initiating, accepting compliments, and asking for help). Through the use of a multiple baseline design the results showed that the participants increased their social interactions with peers. Although the results of the study were positive, the intervention procedures were not thoroughly described, which may make it difficult for future researchers and clinicians to replicate the procedures.

In 2001, Kohler and colleagues evaluated the effects of a "naturalistic" teaching intervention (e.g., using novel materials and incidental teaching) to teach four preschoolers with disabilities social interaction skills (e.g., talking to each other and exchanging materials). The authors also used a multiple baseline design and the results showed clear improvements in social interaction skills following intervention. Although the results were positive, the dependent variables (i.e., social interaction skills) were a constellation of behaviors, so it is not clear exactly which behaviors improved. Despite the limitations with these and other studies that have used single subject designs to evaluate SSGs, researchers have demonstrated that SSGs can be an effective intervention for increasing specific social behavior.

In the last decade the number of studies that have evaluated SSGs using group designs have drastically increased (Kaat and Lecavalier 2014). In 2009, Laugeson and colleagues evaluated the PEERS model for improving social behavior for thirty-three teenagers diagnosed with ASD. The SSG was implemented for 12 weeks, with sessions taking place once a week for 90 min. Each week the authors taught participants a different skill (e.g., choosing appropriate friends, handling teasing, and handling disagreements). The teaching included parent involvement, homework, didactic lessons, and demonstrations. The authors used a group design with a treatment group and a control group (delayed treatment group) to evaluate improvements on various assessments completed by parents and teachers. The results showed that the PEERS model was effective for improving social behavior.

In 2012, Laugeson and colleagues replicated their previous research when they evaluated the UCLA PEERS Program for 28 middle and high school students diagnosed with ASDs. The study consisted of fourteen 90 min sessions following the same procedures and teaching similar skills as in the 2009 study. The results showed that participants made significant gains in social behavior and maintained these gains 14 weeks after intervention had ended. Although the results of these studies are positive, there were a few areas that should be addressed in future research. These areas include implementing a SSG in a more individualized manner and using a blind evaluator to measure behavior change, as opposed to only relying on parents, whose reports may be biased.

While the results of research using both single-subject and group designs have indicated that participants improved in their use of social skills in at least some contexts, there have been limitations within this research which have been described in meta-analysis and review papers (e.g., Cappadocia and Weiss 2011; Kaat and Lecavalier 2014; Rao et al. 2008; Reichow and Volkmar 2010; White et al. 2007). In 2007, White and colleagues evaluated 14 studies that implemented SSGs for individuals diagnosed with ASD. The analysis resulted in the authors concluding that several limitations, within the then current research, should be addressed, including use of control groups, multiple informants completing evaluations, and blind evaluators to conduct independent evaluations. Additionally, the results showed that all studies included participants that were above the age of 6 years-old.

Rao and colleagues (2008) conducted another review of ten studies that implemented SSGs for individuals diagnosed with ASD. The authors concluded that "despite its widespread clinical use, empirical support for social skills training (SST) programs for children with AS/HFA is in its infancy" (p. 359). The authors stated that future researchers should utilize group designs with control groups, participants should be evaluated by blind observers, and researchers should evaluate generalization and maintenance measures. In 2010, Reichow and Volkmar conducted a meta-analysis of sixty-six studies across a variety of social skills interventions, including SSGs. The results of the analysis showed that in some studies maintenance results were poor, SSGs were evaluated as a component of a larger teaching package, and that the mean age of participants in SSGs was 10 years. Thus, future research should evaluate SSGs in isolation and with a younger age group, as well as evaluate long term maintenance.

In a more recent review Kaat and Lecavalier (2014) evaluated 48 studies that implemented SSGs for individuals diagnosed with ASD. In this review, only two studies included participants younger than 6 years of age, the treatment intensity ranged from less than 7 h to 3 years, and only six studies confirmed a diagnosis of ASD through general assessment procedures. Thus, the authors concluded that more work is necessary before firm conclusions regarding the efficacy of SSGs can be made. The authors also recommended that future researchers better describe participants, develop an assessment battery to measure change, and collect more maintenance data.

The reviews of SSGs show a clear need for more studies evaulating SSGs. These studies should include a battery of assessments, better descriptions of participants, evaluation of SSGs for a younger population, the use of randomized control trials, measurement of participants' social behavior via multiple sources, use of blind evaluation, assessment of generalization of skills taught to more natural environments, and assessment of long term maintenance of each participants' social behavior. This study attempts to address these limitations by evaluating a behaviorally based 16-week long SSG for fifteen students under 6 years of age who were diagnosed with ASD. Within this study we utilized a randomized control trial with a waitlist control group, included a battery of assessments to evaluate social behavior, evaluated participants' behavior using three different sources, evaluated each participant's social behavior by a blind assessor, evaluated each participant's social behavior in naturalistic settings, and evaluated each participant's long term maintenance (i.e., 16 and 32 weeks following intervention).

Methods

Recruitment and Inclusion Criterion

Participants were recruited through emails to parents and professionals in the field who have attended previous workshops (approximately 500 local parents and professionals), various announcements on the internet, and through personal contact with professionals not associated with the study. A 20 min screening interview was conducted to determine if potential participants met the inclusion criterion for the SSG, including: (a) no previous history of intervention from the agency conducting the study; (b) a formal and independent diagnosis of ASD using standardized assessments including the Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview™ (ADI™-R); (c) at least 3 years old and less than 7 years old; (d) a full scale IQ score of 80 or above; (e) social deficits displayed on the Social Responsiveness Scale (SRS; Costantino and Gruber 2005) and Social Scale Inventory System (SSiS; Gresham and Elliot 2008); and (f) and displayed average language capabilities in both expressive and receptive language as measured by standardized tests.

Participants

Sixteen children out of 25 respondents met the inclusion criterion and were randomly assigned to the immediate treatment group (referred to as Group A) or the delayed treatment group (referred to as Group B). One of the participants in Group B did not attend the observational sessions (see below) prior to intervention and therefore the study evaluated 15 children (i.e., eight assigned to Group A and seven assigned to Group B). The researchers evaluated each of the 15 participants using the Gilliam Autism Rating Scale-Second Edition (GARS-2) to assess if the scores were consistent with the independent diagnosis of ASD (i.e., a cutoff Autism Index score of at least 70). Table 1 provides demographic information of each participant's age and a battery of assessments across multiple domains (i.e., cognitive, adaptive, social, and language). According to a One-Way ANOVA there was no significant difference between the two groups on the various demographic characteristics (p ranging from .44 to .93).

Table 1 Participant demographic

Domain	Group A	Group B	p value
Number of participants	8	7	
Average age (months)	55	58	.56 (ns)
Average full scale IQ score	101.4	105.7	.45 (ns)
Average vineland adaptive score	83.9	82.9	.92 (ns)
Average expressive one word standard score	108.8	109.1	.93 (ns)
Average peabody picture vocabulary standard score	104.2	108.6	.44 (ns)

A One-Way Anova was conducted to determine if there were differences between the groups in term of Age, cognitive assessments, adaptive behavior assessments, and language assessments

Staff

The entire project was overseen by the lead researcher, who was in charge of recruiting potential participants, helping determine skills to teach, data collection, data analysis, and interacting with parents. Group A and Group B were overseen by two lead teachers (same for both groups) who had 5 years of experience working with individuals diagnosed with ASD and who had led SSGs previously. The lead teachers were responsible for implementing intervention, determining curriculum targets, collecting data, and communicating with parents. There was one support teacher present during each session for Group A and one support teacher present during each session for Group B. Both support teachers had at least 2 years of experience working with individuals diagnosed with ASD. The support teacher was responsible for helping with teaching, managing disruptive behaviors, and collecting data. Finally, a blind evaluator measured each participant's social behavior utilizing formal assessments during group probes and community probes (described below). The blind evaluator had over 10 years of experience working with individuals diagnosed with ASD, had implemented SSGs previously, had previous experience implementing the assessments, and was a board certified behavior analyst.

Setting

Group probes and the SSG (described below) took place in an area that was designed to resemble a large kindergarten classroom within a private clinic that provides behavioral intervention for individuals diagnosed with ASD. The classroom had multiple areas including a large group center, a pretend play center, an indoor play center, and two small group centers. Community probes were conducted in each child's natural environment.

Design

We utilized a pretest–posttest randomized experiment with a wait-list control group design (Campbell and Stanley 1963) to evaluate the effects of the behaviorally based SSG. There were four different assessment periods (i.e., T1, T2, T3, and T4), with Group A receiving 32 sessions of SSG treatment between T1 and T2 and Group B receiving 32 sessions of SSG treatment between T2 and T3. No project-provided treatment occurred between T3 and T4. Thus, the internal validity was strongest for results between T1 and T2, as it was at this period that a randomized control trial design could be used to control for common threats to internal validity such as maturation and history. Although T3 and T4 data did not afford comparisons between groups that are best suited to inferring a treatment effect on social skills measured several weeks after the end of treatment, they did afford as assessment of the stability of scores between each group's post-treatment period and a follow-up. Such stability would provide the basis for future, more internally-valid tests of the hypothesis that the SSG might influence *maintained* social skills.

Procedure Overview

During each assessment period, there were two types of observation sessions: group probes and community probes. There were two, 2 h group probe sessions and one, 20 min community probe. The group probe sessions were semi-structured group sessions that were similar to the SSG sessions with regard to activity type and group composition and were run by the SSG teachers and a support teacher. Importantly, even the group probe sessions differed from the SSG sessions in that the SSG teachers did not use teaching techniques such as prompts or reinforcement for desired social behaviors. Importantly, the community probes were left to vary in setting, structure, activity, materials, location. The adults that might be present during the community probes were never the SSG teachers and varied greatly in how they interacted with the participants. In no case, did the adults systematically use the teaching techniques used in the SSG sessions during the group or community probes. Additionally, the activities, materials, and potential social interactors in the community probes differed from those in the SSG sessions. Thus, the community probes afforded a context in which between-group differences in T1 and T2 changes due to SSG could be said to be *generalized* changes.

Group Probe Sessions and Community Probes

Group Probe Sessions

During each assessment period, there were a total of two group probe sessions per group (four sessions total). Each session lasted 2 h, for a total of 4 h of observation per group per assessment period. Participants in Group A, the ones who received the SSG between T1 and T2, were assessed together at all periods. Similarly, the participants in Group B were assessed together at all periods. In addition to the SSG teachers and support teachers, the lead researcher and the blind evaluator were present during observation sessions, but only observed.

Each group probe session was conducted using the same general schedule. First, the participants were provided with a free play period of time where two or more games or pretend play items were present. Second, the teachers implemented large group instruction where they read a book and asked participants to follow simple instructions. Third,

there was an indoor game period where teachers set up various structured indoor games (e.g., musical chairs). This was followed by an outdoor period at a local park, which consisted of structured games and an unstructured period of time at the playground. After the outdoor period there was another unstructured indoor free-play time which was followed by a final large group instruction.

Except for the types of instructional statements needed to set up the activity, teachers did not provide any prompts to engage in appropriate behaviors or any feedback based upon participant repsonding (neither praise nor corrective feedback). In addition, they did not provide any instructions regarding social behavior, classroom readiness skills, or compliance. Finally, the teachers did not redirect or block any non-dangerous or non-harmful aberrant behavior. If a participant displayed dangerous or harmful behavior, the teachers physically redirected without commenting.

Community Probes

During each assessment period, in addition to observation of the participants in the group setting, the blind observer also observed each participant individually in a community setting. Community probes were 20 min observations that were conducted in environments where the participant spent most of his or her day interacting with peers within her or his community (e.g., school). Whenever possible the community probe was conducted at school during a period that afforded numerous opportunities for social interactions (e.g., recess, free-play, or lunch). For several participants the school did not consent to have an observer present for various reasons, most commonly because there was a pending individualized education plan (IEP) due process litigation. When school was not an option, the community probe occurred in a community setting such as the park where the participant and other children would be present. Thus, the location for the generalization session was at the participant's school whenever possible or in another social setting within the community. During community probes, the blind evaluator did not direct the activities or interactions that occurred. In fact, the degree of structure of interactions, the number of potential social interactors, and the types of activities were free to vary across participants. Because the community probes were left to vary as they would in the natural environment, it is likely that the people, activities, materials, and interaction style of the potential social interactors were quite different than those used in SSG sessions. However, no formal assessment of the degree to which this occurred is available. Considering that the settings of the community probes were also different from the SSG sessions, these sessions afforded tests of generalization across the dimensions of location, activity, materials, person, and interaction style.

Dependent Measures

The four different formal assessments used were the Social Skills Improvement System (SSiS; Gresham and Elliot 2008), Social Responsiveness Scale (SRS; Costantino and Gruber 2005), Walker–McConnell Scale of Social Competence and School (WM; Walker and McConnell 1988) Adjustment, and Aberrant Behavior Checklist (ABC; Aman and Singh 1986). The Social Skills Improvement System and the Walker–McConnell Scale of Social Competence and School Adjustment assess positive social behavior and are constructed in a manner in which high scores are adaptive. In contrast, the Social Responsiveness Scale and the Aberrant Behavior Checklist assess autism social symptoms or aberrant behavior generally associated with autism and are constructed in a manner in which high scores are maladaptive. Within each pair of instruments, while there is some overlap in the type of behaviors measured by each of these instruments, each instrument has items that do not appear in the other instruments and use different wording for similar constructs. We elected to use two instruments for each major construct (social skills and autism-related symptoms) to improve the breadth of assessment for each construct and to provide a comprehensive view of the participants' social functioning.

In contemporary randomized control trials, researchers distinguish between primary and secondary dependent variables. The primary dependent variable is the one on which primary interpretations are based. In this case, we used the scores that the blind evaluator reported because those scores were least likely to be influenced by knowledge of which participants received the SSG at T2 and T3. The secondary dependent variables were scores from the lead researcher and the collaborative scoring of the two SSG teachers. The blind evaluator filled out the four instruments immediately after the community probe, but her scores reflected the participant's performance during group and community probes. In contrast, the lead researcher and the SSG teachers filled out their instruments immediately after the group probe sessions. That is, the lead researcher and SSG teachers did not observe the community probes.

Social Skills Improvement System (SSiS)

The first assessment that was completed by the three evaluators was the SSiS (Gresham and Elliot 2008). The three evaluators used the teacher form and answered 46 different questions about each participant's social behavior. The standard scores (i.e., those for which raw scores were converted to a scale that enabled comparison with informant-reports on a normative sample) were used in subsequent analyses.

Social Responsiveness Scale (SRS)

The second assessment that was completed by the evaluators was the SRS (Costantino and Gruber 2005). The evaluators used the teacher form and either completed the preschool version or the elementary version dependent on the participant's age. The three evaluators answered 65 different questions about a participant's specific social behavior. Standard scores were used in subsequent analyses.

Walker–McConnell Scale of Social Competence and School Adjustment (WM)

The third assessment that was completed by the three evaluators was the WM (Walker and McConnell 1988). The three evaluators used the elementary version and answered 43 questions about social behavior and school readiness behavior. Standard scores were used in subsequent analyses.

Aberrant Behavior Checklist (ABC)

The final assessment that was completed by all three evaluators was the ABC (Aman and Singh 1986). The three evaluators used the community version and filled out all 58 questions about specific aberrant behaviors. The raw scores, consisting of the sum of points from the 58 questions, were used in subsequent analyses because the scale does not have available conversion tables for standardized scores.

There were two dependent variables: (a) an aggregate of the two social skills instruments and (b) an aggregate of the social symptoms instruments. Aggregates (i.e., average z-transformed scores) were used to reduce the number of significance tests and strengthen the validity of the assessment of each construct (i.e., social skill and social symptoms). More on the details of these aggregates are provided in the "Analysis Plan" and "Results" section.

Social Validity

Each participant's parent filled out an anonymous social validity survey (Wolf 1978). The social validity survey consisted of 13 questions about the parent's satisfaction with the group and progress they felt their son or daughter made within the group. Each question used a 7 point Likert scale. For questions that asked about satisfaction parents could provide the following levels of satisfaction: very dissatisfied (1), dissatisfied (2), somewhat dissatisfied (3), neither satisfied nor dissatisfied (4), somewhat satisfied (5), satisfied (6), and very satisfied (7). For questions that asked about improvement parents could provide the following response: great decline (1), decline (2), slight decline (3), neither decline or improvement (4), slight improvement (5), improvement (6), or great improvement (7).

Social Skills Group

The SSG used a behavioral framework in which only evidence-based practices were implemented and all procedures were based upon the principles of ABA. The teachers of the SSG used a progressive model of ABA (Leaf et al. 2016a, c) where the range of ABA based procedures (e.g., discrete trial teaching, shaping, cool versus not cool, role play) that constituted the treatment package were individualized to meet the needs of each participant. Though the term "progressive" is a relatively new theoretical topic (e.g., Leaf et al. 2016a, c) the delivery of this type of model within the field of ABA has been occurring for over 40 years (e.g., Lovaas 1987). The progressive model allowed the teachers greater discretion as compared to strict adherence to protocols or manuals. The teachers used a structured yet flexible approach in teaching and made in-the-moment assessments of what and how best to teach, which is consistent with historical traditions of ABA and has been evaluated in controlled studies (Leaf et al. 2015, 2016b; Soluaga et al. 2008). The decisions made were based on ongoing assessments and constant data analysis. The goal was to maximize the amount of instruction that was provided within the 2 h and minimize any downtime.

Curriculum

Within this study no single, specific curriculum was followed. Instead, the teachers individualized each session's curriculum based upon multiple factors. These factors included: (1) the student's deficits; (2) the group deficits as a whole; (3) deficits identified on the various assessments; and (4) parent concerns. The SSG teachers drew upon a variety of different curriculum books (e.g., Dowd et al. 1994; Laugeson 2014; Taubman et al. 2011), and, at times, the lead teachers created their own curricular targets. There were over 90 targeted social behaviors taught to each group (contact lead author for complete list of skills and when they were taught). Appendix A provides a representative example of the skills that were taught, what procedures were implemented to teach those skills, and the intensity with which they were taught across the 32 sessions for Group A. Appendix B provides the same information for Group B.

Teaching Procedures

A variety of teaching procedures were implemented, all within a progressive ABA framework (Leaf et al. 2016a, c). In addition to listing some of the skills that were taught to

the groups, the Appendices also indicate what procedure(s) were used to teach the various skills taught in the group.

One of the procedures implemented was group discrete trial teaching (e.g., Leaf et al. 2013). Group discrete trial teaching consisted of the teacher providing an instruction, waiting briefly for the participant(s) to respond, and the teacher providing either reinforcement or corrective feedback based upon the participant(s) response. The teachers utilized flexible prompt fading during all teaching trials (Soluaga et al. 2008). The teachers implemented either sequential discrete trials, (each participant had to respond one at a time), overlapping (participants building responses of one and another's answers) or chorally (the entire group responded simultaneously; e.g., Taubman et al. 2011).

A second procedure that was utilized to teach specific social skills was the cool versus not cool procedure (Leaf et al. 2012) in a group instructional format. The cool versus not cool procedure started with the teacher demonstrating the behavior either the appropriate (cool) or inappropriate (not cool) way in front of the group. Next, the teacher asked the participants to rate if the demonstration was cool or not cool and state why. Correct responses resulted in reinforcement and incorrect responses resulted in corrective feedback. This was followed by each of the participants having the opportunity to role-play the behavior the cool way with one of the teachers.

A third procedure utilized to teach specific social behaviors was the teaching interaction procedure (e.g., Kassardjian et al. 2014). The teaching interaction procedure is a multi-step procedure which includes: (a) labeling and identifying the behavior; (b) providing a meaningful rationale to display the behavior; (c) breaking the skill down into smaller components; (d) teacher demonstration; (e) student role-play; and (f) feedback/reinforcement throughout. The teaching interaction procedure was implemented exclusively in large group instructional format. A fourth procedure that was utilized was incidental teaching (Hart and Risley 1975) which was used to promote interaction and social language. Incidental teaching was used throughout sessions as opportunities occurred. For example, if a participant initiated interest with an item, the teacher would follow up with instructions or teaching on that item or interest. Finally, embedded instructions (e.g., Johnson et al. 2005) were also utilized within the context of games to teach social behaviors.

Reinforcement Procedures

The main group reinforcement system was a level system (Hagopian et al. 2002). The level system was a visual reinforcement system in which each participant had his or her own marker which was moved up and down a chart that was divided into levels. The highest level was "Superkid," directly below "Superkid," was "awesome," followed by "okay," followed by "warning," and at the bottom was "miss a fun activity." Each participant had a clip with her or his name. Within each level there were different sublevels so the SSG teacher could move a participant's clip up or down either an entire level or move up or down within a certain level.

There were no set rules of why or when a participant's clip should be moved up or down on the level system, but it was generally contingent upon the current targets as well as each individual's overall social behavior. In general, the SSG teachers would move up a participant's clip for displaying appropriate social behavior, not engaging in disruptive behavior, and for responding correctly. In general, the SSG teachers could move a participant's clip down for not engaging in appropriate behavior or displaying inappropriate behavior.

The participants' clips started each session on "okay" and were moved up and down the chart throughout the session. At any point during a session, if a participant reached "miss a fun activity" he or she would be placed in non-exclusionary time-out (Foxx and Shapiro 1978) and had to watch his or her peers play a fun game (e.g., contingent observation; White and Bailey 1990). If the participant remained calm during this non-exclusionary time out, then they were moved up to "warning" and were placed back in the group. If the participant was not calm, he or she remained in the non-exclusionary time out until he or she became calm. At the end of each session there was a cash-in period. Participants who were on "Superkid" at the end of the session had the opportunity to select an item/toy from a treasure chest to take home. Items were small and included such things as bouncy balls, "flarp™" (a slime substance that makes the sound of flatulence), toy cars, and superhero toys (range $0.33–$3.00 per item). Participants who were not on "Superkid" did not have the opportunity to go to the treasure chest that session. On the final two sessions the cool chart was removed from the group to help participants transition out of the group successfully.

For Group A, the average frequency of reinforcement (i.e., moving up on chart) across all participants was 48.7 (range 0–91 per session) and punishment (i.e., moving down on chart) was 10.1 (range 0–25 per session). The daily ratio of reinforcement to punishment was, on average, 7.3:1. For Group B, the average frequency of reinforcement (i.e., moving up on chart) across all participants per session was 41.2 (range 0–92 per session) and punishment (i.e., moving down on chart) was 9.9 (range 0–29 per session). The daily ratio of reinforcement to punishment was, on average, 5.5:1.

In addition to the level system the SSG teachers introduced a non-contingent group reinforcement system called "Puck the Penguin." Puck the Penguin (named after the

SSG called the Penguins) was a small stuffed animal that each participant could take home for a 1-week period. Accompanying the stuffed animal was a small journal for the participants to write about what they did with Puck while at home. At the end of the week the participant who took Puck home got to present in front of the group about his or her experiences with Puck. Each participant had only one opportunity to take Puck home.

General Structure

Within each group, the teachers had the flexibility to adjust the amount of teaching time for each activity. This adjustment occurred if the lead teachers assessed that more teaching time was needed for a participant(s) to learn a targeted behavior. In general, the group started with a free play period at which participants could interact with dramatic play (e.g., dress up), rule-governed games, or blocks. During this period, the teachers implemented incidental teaching, shaping, and flexible prompt fading. This was followed by the opening circle where the teacher took roll and worked on basic social skills (e.g., observational learning, joint attention, and attending). During this instructional period the teachers implemented shaping, group discrete trial teaching, and the cool versus not cool social discrimination procedure. Third, there was a structured game time at which the participants engaged in multiple structured games (e.g., fruit salad, musical chairs, sleeping game; see Leaf et al. 2016a), within which the teachers embedded specific social behaviors to target. During this period the teachers implemented the cool versus not cool procedure and shaping. This was followed by an outdoor period where the teachers taught structured games (e.g., red light, green light) as well as allowed free play for the participants. During this period the teachers implemented the cool versus not cool procedure, modeling, group discrete trial teaching, and the teaching interaction procedure. Next, there was another large circle where the teachers taught specific social behaviors (e.g., talking to a friend) followed by another round of structured games. During this period the teachers implemented group discrete trial teaching, the cool versus not cool procedure, and the teaching interaction procedure. The group concluded with a closing circle where participants could access their earned reinforcer for the day. During this period the teachers implemented group discrete trial teaching.

Parent Involvement

Parents were welcome to observe all SSG sessions, were debriefed at the end of each session, were able to set up meetings with either the researcher and/or the SSG teachers, were able to invite other family members or professionals to the SSG, and received a bi-monthly letter about the SSG and their son or daughter's progress (contact lead author for example of the letter).

Analysis Plan

Because of the large pool of data to be analyzed, two approaches were taken to protect against type I errors due to multiple significance testing. First, when justified by average within-period correlation between instruments that presumably measured the same construct, we computed the within-individual average of the z-transformed scores from the two measures to form a single aggregate variable. A correlation greater than .50 was deemed sufficient to allow aggregation. Standard or raw scores were transformed to z scores to put them on the same scale. A z-score is computed by the following formula: (mean − observed score)/SD. Thus, z scores have a mean of 0 and a standard deviation of 1. The mean and SD used to compute these were the grand mean and grand SD at T1. This results in a mean of 0 and SD of 1 at T1. Using the T1 mean and SD enables the average of the z-scores to show changes over time.

The SSIS and Walker–McConnell had an across-period mean correlation (i.e., *r*) of .85, .88, and .92, for the SSG teachers, the lead researcher, and the blind evaluator, respectively, with between-scale correlations within observer never falling below our criterion level of .50. For both of these measures, higher scores reflected more adaptive performance. Thus, three within-observer aggregate variables were computed from the SSiS and Walker scores.

The SRS and ABC had a correlation of .67 for the blind evaluator; again, with no period's correlation falling below our criterion level. For both instruments, lower scores reflected adaptive behavior. Thus an aggregate variable was computed from the blind evaluator's SRS and ABC scores. However, computing an aggregate for the social symptoms was not warranted for the SSG teachers or lead researcher because at least one period's correlation between assessments fell below our criterion for each of these observer types. Thus, social symptoms per SSG teacher and lead researchers was quantified at the single instrument level.

In summary, the two primary dependent variables were the blind evaluator's social skill aggregate and the social symptoms aggregate variables. The six secondary dependent variables were (a) the SSG teacher's social skill aggregate variables, (b) the lead researcher's social skill aggregate variable, (c) the SSG teachers' SRS, (d) the SSG teachers' ABC, (e) the lead researcher's SRS, and (f) the lead researcher's ABC. In total, there were eight dependent variables analyzed. Had we not used aggregate variables, there would have been 11 dependent variables.

The second step taken to guard against Type I errors was to use a combination of omnibus statistical testing [i.e.,

mixed Multivariate Analysis of Variance (mMANOVA) and Multivariate Analysis of Covariance (MANCOVA)] and alpha levels that were adjusted for the number of significance tests performed. Specifically, mMANOVAs were used to test the Time × Group, and Time × Group × Dependent Variable interactions for the set of the dependent variables that positively correlated with each other but negatively correlated with the other set of dependent variables. We expected Time × Group interactions that were greater for the non-blind than for the blind observers. When either of the above interactions were significant, figures were visually inspected to determine if our prediction that the between-group difference mainly occurred at Time 2. If the significance test and the visual examination confirmed predictions, we used MANCOVA to control for T1 scores of the dependent variable while testing the between-group difference on the T2 scores. The alphas for all significance tests were adjusted for the number of dependent variables in each mMANOVA or MANCOVA using the Benjamini–Hochberg (1995) method. The size of the set of dependent variables defining a family of comparisons was defined by the number of dependent variables that intercorrelated in the same direction (e.g., positive). We expected the three adaptive aggregates to positively correlate; the five social symptoms variables to positively correlate; and the two sets of dependent variables to negatively correlate with each other. Thus, we expected to carry out two mMANOVAs and two MANCOVAs. Tests of sphericity (mMANOVA) and of homogeneity of slope (MANCOVAs) were planned.

Results

Preliminary Results

As a context, the component variables' means and SD for the primary dependent variable aggregates are given by Time in Table 2. Table 3 indicates the means, SDs, and effect sizes for the between-group difference on the eight dependent variables at T1 to evaluate how well randomization produced comparable groups at T1. While none of the effect sizes were significant, six of the eight between-group differences at T1 were over the threshold that the Institute for Educational Sciences recommends (|.25|) (What Works Clearing House 2016). Thus, we statistically controlled the T1 scores when testing the treatment effect at T2 to improve the accuracy of the effect size estimate and add rigorous protection against the primary alternative explanations for T2 differences: between-group differences at T1 on the dependent variables.

Table 4 provides the intercorrelation of the eight dependent variables, which was the basis for selecting

Table 2 The mean and standard deviation for each assessment conducted by the blind evaluator across the different time periods

Assessment	Group	T1	T2	T3	T4
SSIS	Group A	60 (9.4)	91.5 (12.9)	98.6 (10.4)	106.2 (11.6)
	Group B	58.8 (11.6)	63.4 (12.7)	100.4 (8.7)	104.4 (15.6)
SRS	Group A	67.6 (5.4)	49.5 (3.2)	49.5 (5.3)	50.4 (3.5)
	Group B	68.4 (8.7)	66.4 (9.1)	47.6 (4.4)	52.3 (5.3)
WM	Group A	67 (9.9)	96.8 (10.2)	101.3 (10.3)	109.1 (9.2)
	Group B	68.9 (12.5)	72.9 (17.2)	102.9 (10.2)	107 (12)
ABC	Group A	14.8 (16.5)	6.9 (8.5)	3.4 (4.1)	5.4 (5.7)
	Group B	26.3 (26.5)	18.9 (13.6)	4.6 (3.6)	5.6 (6.8)

Number in parentheses represents the standard deviation

Table 3 Means, SDs, and effect size of between-group difference at Time 1 on dependent variables

Variable	Group A (early treatment)		Group B (late treatment)		Cohen's d for between-group mean difference
	M	SD	M	SD	
Blind social skills aggregate	−.01	.90	.01	1.15	−.02 (ns)
Blind symptom aggregate	.33	.88	−.37	1.02	.74 (ns)
SSG teachers' social skills aggregate	.13	.46	−.15	.54	.56 (ns)
Researcher social skills aggregate	−.15	.62	.17	1.20	−.34 (ns)
SSG teachers' ABC	39.1	26	58	35	−.58 (ns)
SSG teachers' SRS	68.4	10.6	72.3	15.4	−.28 (ns)
Lead researcher's ABC	33.4	27.3	44.9	26.8	−.40 (ns)
Lead researcher's SRS	72.2	3.9	73.4	10.4	−.15 (ns)

Table 4 Zero-order Pearson's product moment correlation coefficients for pairs of dependent variables

	Social skills aggregate from blind evaluator	Social skills aggregate from the SSG teachers	Social skills aggregate from the lead researcher	Social symptom aggregate from the blind evaluator	ABC from the SSG teachers	SRS from SSG teachers	ABC from lead researcher
Social skills aggregate from the SSG teachers	.703**						
Social skills aggregate from the lead researcher	.622*	.780**					
Social symptom aggregate from the blind evaluator	−.729**	−.590*	−.472				
ABC from the SSG teachers	−.623*	−.788**	−.549*	.855**			
SRS from SSG teachers	−.580*	−.693**	−.375	.869**	.895**		
ABC from lead researcher	−.752**	−.758**	−.663**	.733**	.825**	.652**	
SRS from lead researcher	−.610*	−.686**	−.667**	.740**	.666**	.738**	.703**

*p < .05, **p < .01 level (2-tailed)

which dependent variables were entered into the same mMANOVA. As expected the three social skills aggregates positively correlated with each other; the five social symptoms variables positively correlated with each other; and the social skill variables negatively correlated with the social symptoms variables. Thus, Group × Time interactions were tested within two mMANOVAs, three and five dependent variables for skill and symptom constructs, respectively.

Primary Results

Social Skill Aggregates

The data in the mMANOVA on the social skill aggregates did not violate the assumption of sphericity. The Group × Time interaction, $F(3,36) = 52.4$, $p < .001$, with 1.0 power, and the Group × Time × Dependent Variable, $F(6,72) = 3.6$, $p = .003$ with .94 power, were both significant. The 3-way interaction means that the degree to which there were between-group differences in change over time varied as a function of the dependent variable. An examination of the plots by dependent variable indicated that the only large between-group difference occurred at Time 2, but a larger difference between group occurred at Time 2 for the two nonblind aggregates. The MANCOVA on the three social skill aggregates at T2 using the T1 measures of the three social skills aggregates as covariates indicated the data met the assumption of homogeneity of slope for all combinations of the Covariate × Group interaction on all three dependent variables. Therefore, these interaction terms were removed from the analysis to improve statistical power. Relative to Group B, the Group A had superior mean social skill aggregates for the blind evaluator, $F(1,10) = 24.5$, $p = .001$, the SSG teachers, $F(1,10) = 29.7$, $p < .001$, and the lead researcher, $F(1,10) = 137.9$, $p < .001$. Table 5 provides the means, SDs, and between-group effect sizes, which were quite large for all three dependent variables.

Social Symptom Dependent Variables

The data in the mMANOVA on the social symptom dependent variables did violate the assumption of sphericity; therefore, the Greenhouse-Geisser correction was used to test significance of the key interaction. The Group × Time interaction, $F(1.4,17.1) = 10.4$, $p < .001$, with .93 power, and the Group × Time × Dependent Variable, $F(13.3,40.1) = 3.8$, $p = .015$ with .80 power, were both significant. The 3-way interaction means that the degree to which there were between-group differences in change over time varied as a function of the dependent variable. An examination of the plots by dependent variable indicated that the only large between-group difference occurred at Time 2, but a larger difference between groups occurred at Time 2 for the lead researcher's SRS. The MANCOVA on the five social symptom dependent variables at T2 using the T1 measures of the five social symptom dependent variables as covariates indicated the data met the assumption of homogeneity of slope for all

Table 5 Adjusted means, SDs and effect size by observer on the dependent variables at Time 2, controlling for Time 1

	Early M (SD)	Late M (SD)	Between-group *d*
Blind evaluator's social skills aggregate	2.9 (1.1)	.41 (1.4)	2.5**
SSG teachers' social skills aggregate	2.28 (1.3)	−.33 (.81)	2.6**
Researcher's social skills aggregate	2.82 (.80)	−.57 (.68)	4.9**
Blind evaluator's social symptom aggregate	−1.56 (.35)	−.25 (.96)	−2.1**
SSG teacher's ABC	20.5 (23.4)	56.7 (25.4)	−1.33**
SSG teacher's SRS	49.7 (5.6)	74.0 (13.3)	−2.4**
Researcher's SRS	44.9 (4.7)	69.1 (15)	−3.34**
Researcher's ABC	13.1 (15.8)	29 (25.9)	−.74

**p < .01

combinations of the Covariate × Group interaction on all five dependent variables. Therefore, these interaction terms were removed from the analysis to improve statistical power. Relative to Group B, Group A had smaller mean social symptom scores for the blind evaluator, $F(1,15) = 50.4$, $p < .001$; the ABC from the SSG teachers, $F(1,15) = 15.2$, $p = .005$; the SRS from the SSG teachers, $F(1,15) = 35.9$, $p < .001$; and the SRS from the lead researcher, $F(1,15) = 63.9$, $p < .001$. After alpha adjustment, the between-group difference on the ABC from the lead researcher was not significantly different, $F(1,15) = 5.2$, $p > .05$. Table 5 provides the means, SDs, and between-group effect sizes, which were quite large for all three dependent variables. Figure 1 indicates the plots of the two primary dependent variables by time for each group.

Although not the primary way we judge whether treatment is effective, the change on the aggregate variables during the treatment phase (T1–T2 for Group A; T2–T3 for Group B) is indicated in Table 6. Large change occurs for both groups during their treatment phases.

Secondary Results

To evaluate the maintenance of changes, we examined the stability of mean group ratings by the blind evaluator on the four assessment instruments between the end of treatment (T2 for Group A, T3 for Group B) and T4 (see Table 2). Thus the maintenance time period was 32 weeks for Group A and 16 weeks for Group B. For Group A, scores on all four measures remained stable or showed small improvement. For Group B, two of the measures (SSiS and WM) showed slight improvement; the other two measures (SRS and ABC) showed a slight deterioration from T3 to T4, but remained substantially improved from pretreatment (T2) levels. This pattern is evident on Fig. 1.

Social Validity

Table 7 reports the average scores for the social validity questionnaire provided to parents. The scores are reported for parents of Group A, parents of Group B, and a combined average for both groups. Social validity results indicate that parents across both groups were satisfied with the outcomes, the procedures, the skills taught, and with the teachers.

Discussion

The results of this study demonstrated that the implementation of a behaviorally based SSG using a progressive model (Leaf et al. 2016a, c) significantly improved the social behavior for 15 participants diagnosed with ASD. Additionally, the results demonstrated that improvements were only made after the implementation of the SSG; Group A improved their social behavior from T1 to T2 and Group B did not improve their social behavior from T1 to T2. Furthermore, the results demonstrated that participants maintained, and in some cases improved, their social behaviors during long term maintenance assessments. Finally, the study demonstrated that parents of the participants were satisfied with the procedures and the improvements in their son's or daughter's social behavior. These results add and expand upon the current literature for individuals diagnosed with ASD.

The results of this study expand the literature in several ways. Although there have been studies that have used a randomized control trial with a control group (e.g., Laugeson et al. 2009), reviewers have stated that more studies of this type of design are needed in order to have a more complete understanding of how SSGs can contribute to the development of effective treatment for children with ASD (e.g., White et al. 2007; Rao et al. 2008). This study provides important additional data using randomized control

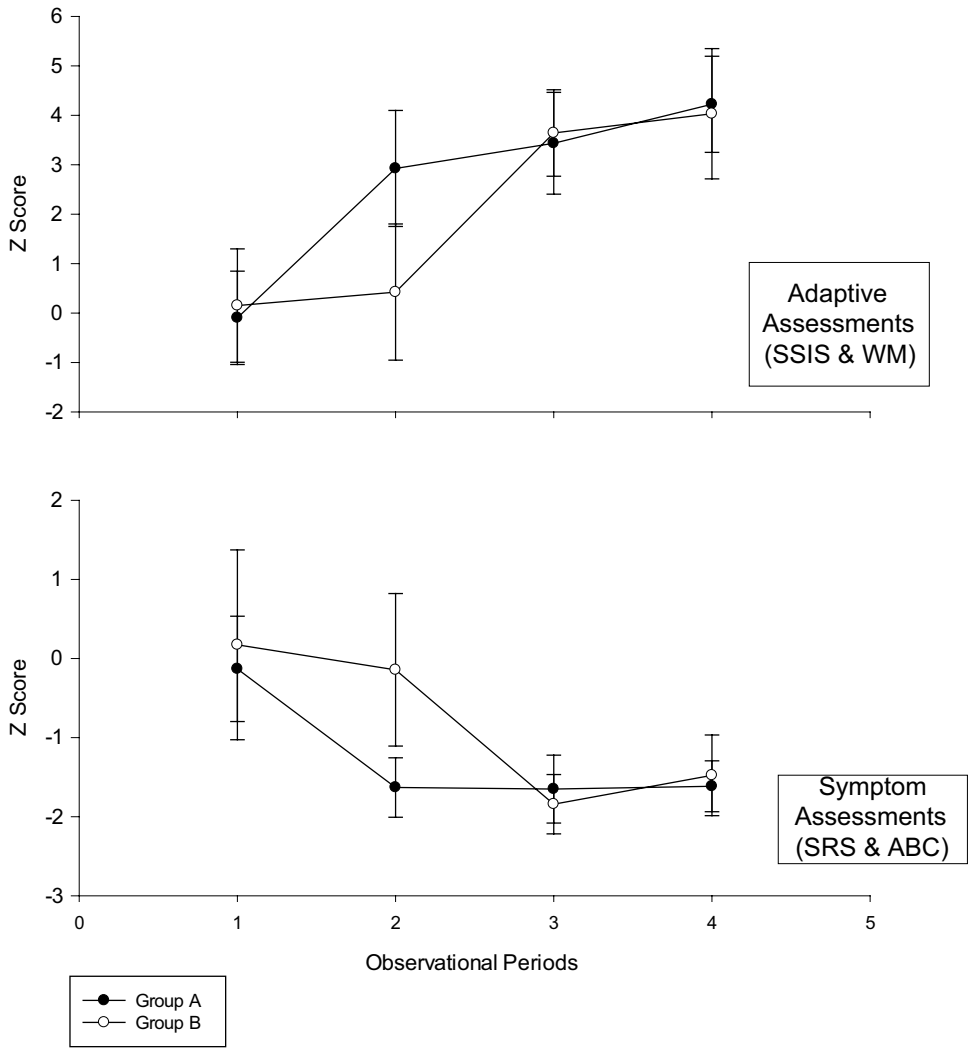

Fig. 1 Means and SD-error bars for the two primary dependent variables by Time and Group

Table 6 Mean and SD of difference between T1 and T2 and within-subjects *d* for change in the treatment phase (T1–T2 for A; T2–T3 for B) by Group

	Early treatment		Later treatment	
	M (SD)	Within-subject *d*	M (SD)	Within-subject *d*
Blind evaluator's social skills aggregate	2.9 (.73)	2.1***	3.2 (.83)	1.6***
SSG teacher's social skills aggregate	2.2 (.98)	1.4***	4.2 (.91)	3.9***
Researcher's social skills aggregate	2.9 (.67)	3.0***	4.5 (.46)	3.6***
Blind evaluator's social symptom aggregate	−1.5 (.57)	−2.1***	−1.7 (.72)	−1.1***

***p < .001

trial methodology. Also, the majority of studies on SSGs for individuals diagnosed with ASD have been conducted with older children (e.g., Kaat and Lecavalier 2014; Reichow and Volkmar 2010). In this study, the average age of the participants was 4 years 10 months; no participant was older than 7 years of age and some participants were only 3 years old at the start of the study. Therefore, this study adds to the current literature on SSGs as it demonstrates that behaviorally based SSGs can be effective for a younger population.

Researchers have also stated a needed area in the research is to evaluate long term maintenance (e.g., Kaat and Lecavalier 2014; Rao et al. 2008). In this study it was found that participants in Group A maintained their skills 16 weeks and 32 weeks after intervention; participants in Group B were maintained their skills 16 weeks after intervention. Therefore, in this study we were able to add to the literature by evaluating long term maintenance and demonstrating that participants maintained their behavior following intervention. Additionally, researchers have

Table 7 Social validity

Question	Average score Group A	Average score Group B	Average score across both Group A and B
Overall how satisfied are you with the social skills group?	6.4	7	6.7
How satisfied are you with your child's ability to learn social skills during the social skills group?	5.8	6.5	6.1
How satisfied are you with your child's ability to learn play skills during the social skills group?	5.8	6.5	6.1
How satisfied are you with your child's ability to learn school readiness skills during the social skills group?	6.2	6.8	6.4
Overall how satisfied are you with the teachers who have run the social skills group?	6.2	7	6.6
How satisfied are you with the teachers' ability to connect with your child?	6.4	6.8	6.6
How satisfied are you with the communication between the teachers and yourself?	6	6.5	6.2
How satisfied are you with the teaching procedures utilized within the social skills group?	6	6.8	6.3
How much improvement do you feel your child has made with his or her social skills?	5.8	6.8	6.3
How much improvement do you feel your child has made with his or her friendship development?	5.6	6.3	5.9
How much improvement do you feel your child has made with his or her school readiness skills?	5.6	6.3	5.9
How much improvement do you feel your child has made in his or her ability to participate in group activities?	5.6	6.5	6
How much do you feel your child has been able to generalize the skills taught in the social skills group to other environments (e.g., school or home)?	5	6	5.4

indicated a need to assess generalization by evaluating participants' social behavior in their naturalistic settings (e.g., school, home, and community; e.g., Rao et al. 2008). In this study, the blind evaluator conducted observations four times in the participants' natural settings (i.e., school, home, or community); these evaluations were included as part of the scoring of each of the four assessments used in the study. Therefore, we were able to evaluate generalization of participants' social behaviors in more natural settings. However, due to logistical constraints we were unable to measure specific social behaviors in participants' naturalistic environments; thus, this type of data is still needed in future research studies.

Perhaps the biggest contribution of this study was the use of a blind evaluator as the main assessor of participants' improvements. Previous reviewers have cited this as one of the biggest limitations in the SSG research (e.g., Rao et al. 2008; White et al. 2007) and a needed component of SSG research. Adding a blind evaluator helps reduce the potential for biased reporting and adds experimental strength to the study. The blind evaluator was not the only professional evaluating participants' progress; progress was also measured by the SSG teachers and researchers. Thus, this study adds to the literature by using multiple evaluators, an area that has been reported as a limitation of the SSG literature (e.g., White et al. 2007). Finally, a battery of assessments was used in this study to evaluate progress, another area of need within the SSG literature (e.g., Kaat and Lecavalier 2014).

A final contribution to the literature is that the teachers in this study used a progressive model of behavioral intervention (e.g. Leaf et al. 2016a, c). In previous research studies on SSGs both the curriculum and procedures were prescribed ahead of time (e.g., Laugeson et al. 2009, 2012). Although strictly adhering to fixed protocols makes it easier to replicate, it restricts clinicians' and researchers' ability to individualize the intervention and to make adjustments to best meet the day by day and minute by minute needs of participants. In 2016, Leaf and colleagues wrote a commentary arguing for the advantages of in-the-moment decision making, flexible prompt fading, and use of a variety of curriculum; all of which have been empirically evaluated (e.g., Leaf et al. 2015, 2016b; Lovaas 1987; Soluaga et al. 2008). In line with this argument, the procedures used in this study were aligned with a progressive model of intervention. A broad range of curricula and teaching procedures were used in the present study, information is contained herein detailing the skills taught, the progression of the skills that were taught, the amount of reinforcement provided for both groups, and the amount of reinforcement and punishment provided for both groups. Thus, this study provides future researchers and clinicians with guidelines of how to implement a progressive behavioral SSG.

Despite the positive outcomes of using a progressive model of ABA (e.g., learner skill acquisition), this model also has some limitations with respect to research. For one, it requires in-the-moment assessments and decisions by the teacher which may lead to one teacher selecting to provide a prompt or teach a skill with a certain procedure and a different teacher selecting different prompts or procedures. This may create difficulties in quantifying treatment decisions and ensuring a high degree of fidelity across teachers. Nonetheless, we attempted to show partial treatment integrity by comparing reinforcement rates across the two groups which showed that they were fairly similar. Although in clinical practice we consider it an advantage to have available a broad range of tools and to switch between them as needed, it is also a limitation of this study, from the standpoint of replicability. However, we provided an analysis of the skills taught, how they were taught in an attempt to describe what occurred for the 64 h of intervention each group received, and provided an analysis of the consequences provided across the two groups. Future researchers may wish to evaluate how to measure treatment fidelity when using this model in the context of a group setting.

There were additional limitations and areas that should be addressed by future researchers. For one, this study was implemented by professionals who had years of experience implementing SSGs; therefore, it is not known what the effects would be when implemented by less experienced teachers and this should be examined by future researchers. At the start of this study we attempted to have parents and school teachers provide us scores on the four assessments; however, we were unable to gather sufficient information from teachers and parents and, therefore, these measures were not reported. Future researchers should attempt to measure participants' change in social behavior from both blind evaluators and parents to see if scores would be aligned. Third, this study implemented the SSG for participants who would be considered "higher" functioning and, therefore, future researchers should evaluate SSGs for lower functioning participants. Fourth, there were only a small number of participants evaluated within this study. Future researchers, may wish to evaluate the procedures on a larger scale. Finally, future researchers should attempt to implement and evaluate SSGs that use a progressive model and continue to operationally define the components of this model of intervention.

Despite the limitations and areas needed for future research this study clearly demonstrated that the implementation of the SSG helped improve the social behaviors for these individuals diagnosed with ASD. The study showed that without intervention participants' social behavior did not improve and that it was not until intervention was implemented that behavior changes were observed. Additionally, this study adds to the current research on SSGs. The results provide guidance to practitioners on how to address the social skills deficits of children with ASD and results such as obtained here could be life altering for individuals diagnosed with ASD.

Acknowledgments We wish to thank Andi Waks, Michele Jessner, Clifford Anderson, Jan Arter, Lupe Ortega, and Andrew McEachin for their help on this project. We also wish to thank the families and the penguins (the children) who participated in this study.

Author Contributions JBL conceived, helped design, and coordinated the implementation of the study, as well as drafted the manuscript; JAL participated in the implementation of the procedures and interpretation and analysis of the data; CM participated in the implementation of the procedures and interpretation and analysis of the data; MT helped design the study and was critical in drafting the manuscript; MOL participated in the acquisition and analysis of the data; NT participated in the acquisition and analysis of the data; DTC participated in the acquisition and analysis of the data; RL helped design the study and was critical in drafting the manuscript; JMC helped design the study and was critical in drafting the manuscript; PY helped in the analysis of the data and interpretation of the data and drafting the manuscript.

Compliance with Ethical Standards

Conflict of Interest This study was partially funded by a grant received from the Organization for Autism Research. This grant was directly provided to the first author, fourth author, fifth author, eighth author, and ninth author. The fifth author received an honorarium for her involvement. There are no conflicts of interest for the remaining authors. Additionally, the fourth, eight, and ninth author have commercial products for procedures used within this study and own an agency that provides behavioral intervention, which include procedures similar to those evaulated here, for individuals diagnosed with ASD.

Ethical Approval All procedures performed in studies involving human participants were in accordance with ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Appendix A

See Table 8.

Table 8 Skills taught in Group A

Skill	Teaching Procedure(s)	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32
Social Discrimination	CNC								
Outdoor Games	S, DTT, MODEL, INC								
Reduction of Aberrant Behavior	S, SR^+								
Structured Games (e.g., Fruit Salad, Mouse Trap, Sleeping Game)	CNC								
Conditioning Peers as Reinforcers	S, FPF								
Positive Affect and Being Silly	S, DTT								
Observational Learning	DTT, INC								
Conditional Instructions	DTT								
Inferences and Predictions	DTT								
Providing Personal Information and Favorites	DTT, FPF								
Attending	DTT, FPF								
General Knowledge and Pop Culture Knowledge	DTT, FPF								
Playing with a Friend	INC, S, MODEL								
Answering and Asking Questions	DTT, INC								
Joint Attention	DTT, S								
Figuring It Out, Trying, and Guessing	S, FPF								
Having Fun with Friends	INC, S, SR^+								
Sitting and Waiting	DTT, SR^+								
Walking in a Line	DTT, FPF								
Winning and Losing Graciously	CNC, TIP								
Conversation	TIP								
Fluency	DTT								
Instructions Through Musical Games	CNC, MODEL, INC								
Joining In	S, SR^+								
Pretend Play	TIP, CNC, MODEL								
Identifying Peers in the Group	DTT								
Social Orientation	S								
Compliance	S, SR^+								
Frustration Tolerance	DTT, MODEL								
Sharing & Turn Taking	CNC, TIP, MODEL								
Contingency Development	S, SR^+								
On Topic Statements	DTT, MODEL, TIP								
Flexibility	S, FPF, SR^+								

Black bars represent skills targeted in 3–4 sessions; gray bars represent skills targeted in 1–2 sessions; white bars represent skills not targeted

S shaping, *FPF* flexible prompt fading, *MODEL* modeling, SR^+ reinforcement, *TIP* teaching interaction procedure, *INC* incidental teaching

Appendix B

See Table 9.

Table 9 Skills taught in Group B

Skill	Teaching Procedures	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32
Figuring It Out, Trying, and Guessing	S, FPF								
Inferences and Predictions	DTT								
Instructions Through Musical Games	CNC, MODEL, INC								
Outdoor Games	S, DTT, MODEL, INC								
Conditional Instructions	DTT								
Playing with a Friend	INC, S, MODEL								
Social Discrimination	CNC								
Providing Personal Information and Favorites	DTT, FPF								
Reduction of Aberrant Behavior	S, SR^+								
Pretend Play	TIP, CNC, MODEL								
Structured Games (e.g., Fruit Salad, Mouse Trap, Sleeping Game)	CNC								
Winning and Losing Graciously	CNC, TIP								
Frustration Tolerance	DTT, MODEL								
Joining In	S, SR^+								
Observational Learning	DTT, INC								
Positive Affect and Being Silly	S, DTT								
Conversation	TIP								
Joint Attention	DTT, S								
Compliance	S, SR^+								
Conditioning Peers as Reinforcers	S, FPF								
Social Orientation	S								
Answering and Asking Questions	DTT, INC								
Flexibility	S, FPF, SR^+								
General Knowledge and Pop Culture Knowledge	DTT, FPF								
Fluency	DTT								
Attending	DTT, FPF								
Having Fun with Friends	INC, S, SR^+								
On Topic Statements	DTT, MODEL, TIP								
Contingency Development	S, SR^+								
Identifying Peers in the Group	DTT								
Walking in a Line	DTT, FPF								
Sharing & Turn Taking	CNC, TIP, MODEL								
Sitting and Waiting	DTT, SR^+								

Black bars represent skills targeted in 3–4 sessions; gray bars represent skills targeted in 1–2 sessions; white bars represent skills not targeted

S shaping, *FPF* flexible prompt fading, *MODEL* modeling, SR^+ reinforcement, *TIP* teaching interaction procedure, *INC* incidental teaching

References

Aman, M. G., & Singh, N. N. (1986). *Aberrant behavior checklist manual*. East Aurora, NY: Slosson Publications.

Barry, T. D., Klinger, L. G., Lee, J. M., Palardy, N., Gilmore, T., & Bodin, D. S. (2003). Examining the effectiveness of an outpatient clinic based social skills group for high-functioning children with autism. *Journal of Autism and Developmental Disorders, 33*, 685–701. doi:10.1023/B:JADD.0000006004.86556.e0.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, 289–300.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Company.

Cappadocia, C. M., & Weiss, J. A. (2011). Review of social skills training groups for youth with Asperger syndrome and high functioning autism. *Research in Autism Spectrum Disorders, 5*, 70–78. doi:10.1016/j.rasd.2010.04.001.

Costantino, J. N., & Gruber, C. P. (2005). *Social Responsiveness Scale (SRS)*. Los Angeles, CA: Western Psychological Services.

Dowd, T., Czyz, J. D., O'Kane, S. E., & Elfson, A. (1994). *Effective skills for child-care workers: A training manual from Boys Town*. Boys Town, NE: Boys Town Press.

Foxx, R. M., & Shapiro, S. T. (1978). The timeout ribbon: A non-exclusionary timeout procedure. *Journal of Applied Behavior Analysis, 11*, 125–136. doi:10.1901/jaba.1978.11-125.

Gresham, F. M., & Elliot, S. N. (2008). *Social skills improvement system rating scales manual*. Minneapolis, MN: NCS Pearson.

Hagopian, L. P., Rush, K. S., Richman, D. M., Kurtz, P. F., Contrucci, S. A., & Crossland, K. (2002). The development and application of individualized level systems for the treatment of severe problem behavior. *Behavior Therapy, 33*, 65–86. doi:10.1016/S0005-7894(02)80006-5.

Hart, B., & Risley, T. R. (1975). Incidental teaching of language in the preschool. *Journal of Applied Behavior Analysis, 8*, 411–420. doi:10.1901/jaba.1975.8-411.

IAN Research Findings. (2011). *Treatment series introduction*. Retrieved from http://www.iancommunity.org/cs/ian_treatment_reports/overview.

Johnson, J. W., McDonnell, J., Holzwarth, V. N., & Hunter, K. (2005). The efficacy of embedded instruction for students with developmental disabilities enrolled in general education classes. *Journal of Positive Behavior Interventions, 6*, 214–227. doi:10.1177/10983007040060040301.

Kaat, A. J., & Lecavalier, L. (2014). Group-based social skills treatment: A methodological review. *Research in Autism Spectrum Disorders, 8*, 15–24. doi:10.1016/j.rasd.2013.10.007.

Kamps, D. M., Leonard, B. R., Vernon, S., Dugan, E. P., Delquadri, J. C., Gershon, B., Wade, L., & Folk, L. (1992). Teaching social skills to students with autism to increase peer interactions in an integrated first-grade classroom. *Journal of Applied Behavior Analysis, 25*, 281–288. doi:10.1901/jaba.1992.25-281.

Kassardjian, A., Leaf, J. B., Ravid, D., Leaf, J. A., Alcalay, A., Dale, S., Tsuji, K., Taubman, M., Leaf, R., McEachin, J., & Oppenheim-Leaf, M. L. (2014). Comparing the teaching interaction procedure to social stories: A replication study. *Journal of Autism and Developmental Disorders, 44*, 2329–2340. doi:10.1007/s10803-014-2103-0.

Kohler, F. W., Anthony, L. J., Steighner, S. A., & Hoyston, M. (2001). Teaching social interaction skills in an integrated preschool: An examination of naturalistic tacts. *Topics in Early Childhood Special Education, 21*, 93–103. doi:10.1177/027112140102100203.

Laugeson, E. A. (2014). *The peers curriculum for school-based professionals: Social skills training for adolescents with autism spectrum disorder*. London: Routledge.

Laugeson, E. A., Frankel, F., Gantman, A., Dillon, A. R., & Mogil, C. (2012). Evidence-based social skills training for adolescents with autism spectrum disorders: The UCLA PEERS program. *Journal of Autism Developmental Disorder, 42*, 1025–1036. doi:10.1007/s10803-011-1339-1.

Laugeson, E. A., Frankel, F., Mogil, C., & Dillon, A. R. (2009). Parent-assisted social skills training to improve friendships in teens with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*, 596–606. doi:10.1007/s10803-008-0664-5.

Leaf, J. A., Leaf, J. B., Milne, C., Townley-Cochran, D., Oppenheim-Leaf, M. L., Cihon, J. H., Taubman, M., McEachin, J., & Leaf, R. (2016a). The effects of the cool versus not cool procedure to teach social game play to individuals diagnosed with autism spectrum disorder. *Behavior Analysis in Practice, 9*, 34–49. doi:10.1007/s40617-016-0112-5.

Leaf, J. B., Leaf, R., Alcalay, A., Leaf, J. A., Ravid, D., Dale, S., Kassardjian, A., Tsuji, K., Taubman, M., McEachin, J., & Oppenheim-Leaf, M. (2015). Utility of formal preference assessments for individuals diagnosed with autism spectrum disorder. *Education and Training in Autism and Developmental Disabilities, 50*, 199–212.

Leaf, J. B., Leaf, R., Leaf, J. A., Alcalay, A., Ravid, D., Dale, S., Kassardjian, A., Tsuji, K., Taubman, M., McEachin, J., & Oppenheim-Leaf, M. L. (2016b). Comparing paired-stimulus preference assessments to the in-the-moment reinforcer analysis on skill acquisition: A preliminary analysis. *Focus on Autism and Other Developmental Disabilities*. doi:10.1177/1088357616645329.

Leaf, J. B., Leaf, R., McEachin, J., Taubman, M., Ala'i-Rosales, S., Ross, R. K., Smith, T., & Weiss, M. J. (2016c). Applied behavior analysis is a science and, therefore, progressive. *Journal of Autism and Developmental Disorders, 46*(2), 720–731. doi:10.1007/s10803-015-2591-6.

Leaf, J. B., Tsuji, K. H., Griggs, B., Edwards, A., Taubman, M., McEachin, J., Leaf, R., & Oppenheim-Leaf, M. L. (2012). Teaching social skills to children with autism using the cool versus not cool procedure. *Education and Training in Autism and Developmental Disabilities, 47*, 165–175.

Leaf, J. B., Tsuji, K. H., Lentell, A. E., Dale, S. E., Kassardjian, A., Taubman, M., McEachin, J., Leaf, R., & Oppenheim-Leaf, M. L. (2013). A comparison of discrete trial teaching implemented in a one-to-one instructional format and in a group instructional format. *Behavioral Interventions, 28*, 82–106. doi:10.1002/bin.1357.

Lovaas, I. O. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3–9. doi:10.1037/0022-006X.55.1.3.

Matson, J. L., Matson, M. L., & Rivet, T. T. (2007). Social-skills treatments for children with autism spectrum disorders: An overview. *Behavior Modification, 31*, 682–707. doi:10.1177/0145445507301650.

Rao, P. A., Beidel, D. C., & Murray, M. J. (2008). Social skills interventions for children with aspergers syndrome or high functioning autism: A review and recommendations. *Journal of Autism and Developmental Disorders, 38*, 353–361. doi:10.1007/s10803-007-0402-4.

Reichow, B., & Volkmar, F. R. (2010). Social skills interventions for individuals with autism: Evaluation for evidence-based practices within a best evidence synthesis framework. *Journal of Autism and Developmental Disorders, 40*, 149–166. doi:10.1007/s10803-009-0842-0.

Sartini, E. C., Knight, V. F., & Collins, B. C. (2013). Ten guidelines to facilitate social groups for students with complex special needs. *Teaching Exceptional Children, 45*, 54–62. doi:10.1177/004005991304500306.

Soluaga, D., Leaf, J. B., Taubman, M., McEachin, J., & Leaf, R. B. (2008). A comparison of flexible prompt fading and constant time delay for five children with autism. *Research in Autism Spectrum Disorders, 2*, 753–765. doi:10.1016/j.rasd.2008.03.005.

Taubman, M., Leaf, R., & McEachin, J. (2011). *Crafting connections: Contemporary applied behavior analysis for enriching the social lives of persons with autism spectrum disorder*. New York, NY: DRL Books.

Walker, H. M., & McConnell, S. R. (1988). *Walker–McConnell scale of social competence and school adjustment: A social skills rating scale for teachers*. Austin, TX: Pro-Ed.

What Works Clearing House. (2016). *Procedures and standards handbook, 3rd version*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_draft_standards_handbook.pdf.

White, A. G., & Bailey, J. S. (1990). Reducing disruptive behaviors of elementary physical education students with sit and watch. *Journal of Applied Behavior Analysis, 23*(3), 353–359. doi:10.1901/jaba.1990.23-353.

White, S. W., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism Developmental Disorders, 37*, 1858–1868. doi:10.1007/s10803-006-0320-x.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heat1. *Journal of Applied Behavior Analysis, 11*, 203–214. doi:10.1901/jaba.1978.11-203.