

University of Nevada, Reno

**The Assessment and Treatment of Conversational Skills for Individuals with
Autism Spectrum Disorder: A Replication and Extension**

A thesis in partial fulfillment of the requirements for the degree Master of Science in
Psychology

by
Kathleen Wiley

Bethany Contreras, Ph.D./Thesis advisor

December 2022



THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

KATHLEEN WILEY

entitled

**The Assessment and Treatment of Conversational Skills for
Individuals with Autism Spectrum Disorder: A Replication
and Extension**

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN PSYCHOLOGY

Bethany Contreras, Ph.D.

Advisor

Matthew Lemon, Ph.D.

Committee Member

Maryann Demchak, Ph.D.

Graduate School Representative

Markus Kemmelmeier, Ph.D., Dean

Graduate School

December, 2022

Abstract

Deficits in social skills are considered a hallmark of autism spectrum disorder (ASD) and have been shown to impact quality of life, but the research on effective social skills interventions for the adolescent population is lacking. Specifically, more research is needed on complex social skills, like conversation skills, and methods that promote generalization. The current study is a replication and extension of a study by Hood and colleagues (2017) that used behavioral skills training and in-vivo training to teach conversation skills to adolescents with autism spectrum disorder, and a semi-structured direct conversation assessment to assess generalization of skills.

Acknowledgements

I'd first like to thank my advisor, Dr. Bethany Contreras, for her ongoing support and guidance throughout this process. Her insight and enthusiasm were invaluable to the completion of this project. I would also like to thank Dr. Larry Williams, who opened the door to the world of behavior science for me. Further, I'd also like to extend a thank you to the Contreras lab and all the research assistants that had hands in this project; it would not have been possible without your help. Finally, I'd like to thank my dad for his love, encouragement, and "I'm proud of you" moments throughout this entire process.

Table of Contents

Abstract	i
Acknowledgements	ii
List of Tables	iv
List of Figures	v
The Assessment and Treatment of Conversational Skills for Individuals with Autism Spectrum Disorder: A Replication and Extension	1
Direct Assessment of Social Skills	2
Treatment of Social Skills	7
Method	12
Participants, Settings, Materials	12
Dependent Variables and Interobserver Agreement	14
Data Collection	14
Interobserver Agreement	15
Experimental Design and Procedures	17
Initial Assessment	18
Baseline and Assessment Probes	19
Behavioral Skills Training	20
In-Vivo Teaching	21
Procedural Integrity	22
Social Validity	24
Affect Data	24
Results and Discussion	26
Results	26
Participant 1	28
Participant 2	32
Discussion	42
Participant 1	42
Participant 2	44
Limitations and Future Research	45
Summary	50
References	51

List of Tables

Table 1. Target behavior operational definitions and measures for each participant

Table 2. Interobserver agreement data Participant 1

Table 3. Interobserver agreement data Participant 2

Table 4. Procedural integrity

Table 5. Pre-treatment social validity Participant 1

Table 6. Post-study social validity Participant 1

Table 7. Pre-treatment social validity Participant 2

List of Figures

Figure 1. Participant 1 results on assessment probes

Figure 2. Participant 1 BST training data target behavior 1

Figure 3. Participant 1 BST training data target behavior 2

Figure 4. Participant 1 affect data

Figure 5. Participant 2 results on assessment probes

Figure 6. Participant 2 BST/in-vivo training data target behavior 1

Figure 7. Participant 2 affect data

Figure 8. Participant 2 indices of displeasure/boredom data

The Assessment and Treatment of Conversational Skills for Individuals with Autism Spectrum Disorder: A Replication and Extension

Deficits in social skills are often considered the "core deficit" or "central feature" of autism spectrum disorder (ASD; Gillis et al., 2011; Reichow & Volkmar, 2010). In a review of the literature, Reichow and Volkmar (2010) noted that a number of studies "identified social difficulties [...] as being the single most powerful predictor of diagnostic status [of autism]" (p. 149). Gaps in the social repertoire have been shown to impede an individual's ability to function adaptively in their environment; for example, people with autism are much more likely to experience difficulty creating meaningful social relationships, which can lead to peer rejection, anxiety, depression, etc. (Gillis et al., 2011).

Social skills are under-researched compared to other aspects of ASD, especially for adolescent and adult populations (Reichow & Volkmar, 2010; Verhoeven et al., 2013; Camargo et al., 2014). While Reichow and Volkmar's review revealed an increasing trend in journal articles on social skills interventions, the research that is available appears to largely focus on interventions for young children with ASD. In their review, Reichow & Volkmar reported that only three of the 66 studies included focused on adolescent and adult participants. While more research has been published since this review, the number of studies including adolescents and adults are still lacking, social skills are socially significant targets for this specific population; social environments and interactions tend to become more complicated in adolescence and into early adulthood, making teens and young adults more likely to be aware of and affected by these types of deficits (Davis et al., 2010; Verhoeven et al., 2013).

The lack of research on social skills for adolescents and young adults may be attributed to the complexity of assessing social deficits and teaching social behaviors (Freeman & Cronin, 2017). Adolescents and adults are exposed to more complex and nuanced social situations as compared to children (Davis et al., 2010). Conversations, for example, are typically composed of multiple verbal and nonverbal speaker and listener skills that vary in complexity (Dotson et al., 2010; Beaulieu et al. 2013; Hood et al., 2017). A robust, socially adept repertoire often requires contextual responding and subtle discriminations in a wide variety of settings. Even if a person is taught a particular skill, they may not be able to read a social situation such that they know when it's most appropriate to use the skill (Davis et al., 2010). Therefore, the assessment and treatment of social skills must be comprehensive and flexible in order to capture the wide variety of issues and deficits (Hood et al., 2017; Freeman & Cronin, 2017). Among individuals and across settings, there is variability regarding what may be considered "socially appropriate" (Freeman & Cronin, 2017). There is even disagreement and inconsistency for how to identify, define, and measure behaviors most valuable to a robust social skills repertoire (Matson & Wilkins, 2007; Veroheven et al., 2013).

Direct Assessment of Social Skills

The development of reliable and valid direct assessments of social skills can help address some of the noted issues. Additionally, more empirically supported direct assessments may further refine the development of interventions by aiding behavior change agents in defining and identifying skills to target for remediation (Gillis et al., 2011). Although there are tools available that are designed to assess social skills, the vast

majority of them are indirect and rely on third-party or self-report (Gillis et al., 2011; Ratto et al., 2011; Veroheven et al., 2013).

However, there are increasing numbers of studies that have used direct assessment methods. For example, Lerman and colleagues (2017) developed a direct assessment for evaluating workplace relevant social skills in adolescents and adults. Eight participants, aged 16-32, participated in the study. Target behaviors were identified through a literature review of studies evaluating and treating job-related behaviors in individuals with developmental disabilities, and performance was assessed by exposing each participant to various conditions (8-15 minutes in length) to see how they would respond. For instance, in one condition, the participant was given a work task within their repertoire that required a specific set of materials. They were purposefully not provided with a complete set of materials to assess whether or not the participant would approach the “supervisor” to ask for the required materials. Participants experienced each of the conditions across several sessions. Observers collected data on specific target behaviors that were objectively defined then directly measured. Data was quantified a variety of ways; the researchers used percentage of opportunities (for the target behavior of making confirming statements), percentage of steps in task analysis completed correctly (for targets asking for help, asking for materials, responding to corrective feedback, notifying supervisor of task completion), and percentage of intervals using 10-s whole interval recording (to track on-task behavior and inappropriate behavior). The results of the study indicated that the assessment was a useful clinical tool to identify workplace social skills to target for intervention.

Veroheven et al. (2013) used a “performance-based measure” called the “Social Skills Performance Assessment” (SSPA) to assess the verbal and nonverbal social skills of 44 adult males (ages 18-35 years) with ASD. Participants were asked to perform in two specific social interactions through two “highly protocolized and standardized role plays of 3 minutes” (p. 2992) with a confederate. The two role-play scenarios included the confederate acting as a new neighbor and then a landlord. Six categories of social skills (interest/disinterest, fluency, clarity, focus, affect, and social appropriateness) were rated on a Likert scale from 1 to 5. Higher scores indicated better demonstration of the social skill.

Another direct assessment of social skills, developed by Ratto et al. (2011), is called the “Contextual Assessment of Social Skills” (CASS). The CASS is a role-play assessment in which participants are exposed to two different three-minute conditions with a confederate. In the first condition, the confederate is instructed to appear bored or disinterested throughout the conversation. In the second condition, the confederate appears interested in the conversation. Conversations were videotaped and trained observers collected data on ten categories of the participant’s verbal and nonverbal behaviors (asking questions, topic changes, vocal expressiveness, gestures, positive affect, posture, kinesic arousal, social anxiety, overall involvement in the conversation, and overall quality of rapport). Observers took frequency data on asking questions and topic changes, and rated all other target behaviors on a Likert scale from 1 (low) to 7 (high). In a study piloting the CASS, the researchers recruited twenty adolescent or young adults with high functioning autism and twenty undergraduate young adult participants to serve as a control. Ratto and colleagues (2011) reported that the control

participants showed “significant change” between conditions in their behavior, while participants with high functioning autism showed little change between conditions, indicating that their assessment procedures were effective in detecting differences in performance across groups and conditions. Their results also allow for a comparison between normative data from typically developing children and children with high functioning autism.

A semi-structured direct assessment, developed by Hood et al., (2017), helped them identify greeting and conversation skills deficits in one child and two adolescents with ASD. The semi-structured assessment was a 45-60 minute conversation with a confederate and multiple embedded “evocative situations” which probed skill competence for a number of conversational behaviors of interest. Target behaviors for the “evocative situations” were operationally defined and selected based on social skills deficits identified in the literature for the Autism population and parent interview prior to the assessments. Examples of “evocative situations” for target social/conversation behaviors included giving a compliment, asking a question, introducing a controversial topic, giving an index of boredom, etc. Each of the ten “evocative situations” were embedded at least three times during the conversation to obtain a sample of the participant’s competency with each specific skill. Participants were all exposed to the same set of evocative situations during the initial assessment. Responses to the evocative situations were either scored as correct or incorrect. Other nonverbal behaviors were also tracked such as smiling, eye contact, and distracting non-vocal behavior.

Custer et al. (2020) modified the assessment protocol used by Hood et al. (2017) to identify conversation skills to target in their study that evaluated the effectiveness of

individualized computer-based instruction and practice with peers. Five participants aged 17-33 were recruited to participate in the study. The assessment lasted 60 minutes and 15 conversation skills were evaluated. Ten of the conversation skills were measured continuously and five were responses to evocative situations embedded into the semi-structured conversation. Data collected was then compared to typical peers to further identify socially inappropriate responses. Custer and colleagues used the results of the assessment to identify three conversation skills to target for each participant, individualized to their own deficits observed in the assessment.

Overall, direct assessments of social skills are less efficient and cost effective than indirect assessments such as questionnaires. However, indirect assessments are limited by the biases of third-party reporters, and there is often poor reliability across reporters (Ratto et al., 2011). Direct assessments give practitioners and researchers an objective sampling of the participant's repertoire. The CASS (Ratto et al., 2010), SSPA (Veroheven et al., 2013), and Lerman and colleagues' (2017) assessments are all examples of highly protocolized role-play direct assessments of social skills. The role-play format is designed to probe performance in very specific social scenarios, so they may not capture a broad picture of the participant's overall repertoire. Many of the role-play direct assessments (CASS and SSPA) also seem to use Likert-scale type ratings as their primary measure, which has similar limitations as indirect assessments in that ratings can be heavily impacted by observer biases. The semi-structured conversation assessment by Hood and colleagues (2017) and modified by Custer et al. (2020) is more time and resource intensive because a wider range of targets are being assessed and each target behavior is objectively defined and directly measured but may also be more

comprehensive and flexible. Using this semi-structured format may better allow clinicians and researchers to assess and treat a wide variety of conversation skills.

Treatment of Social Skills

The social skills interventions for the ASD population with the most empirical support are social skills groups and video modeling (Reichow & Volkmar, 2010). Sansosti and Powell-Smith (2008) evaluated the effects of video modeling and computer-presented social stories on social skills of three children with high functioning ASD. One target behavior was selected for each participant based on direct observation. The video modeling intervention involved participants watching a video of someone modeling the target behavior. One social story for each participant was also created. Data collected during unstructured direct observations of participants showed improved rates of social communication following the implementation of the treatment package. Maintenance of skills was demonstrated for each participant at a two-week follow-up, but generalization of skills was only observed with one participant. One potential limitation of video modeling is the limited demonstration of generalization of skills across studies (Nuernberger et al., 2013). Another limitation is it requires the use of additional, sometimes expensive, materials to create the videos. In addition, creating enough videos that capture a wide range of social situations sufficient for multiple exemplar training may be time and resource intensive.

Most of the reported findings using social skills groups as an intervention report positive outcomes. For example, Tse et al. (2007) used a social skills training group to teach social skills to adolescents with high functioning ASD. The effects of their intervention were evaluated using questionnaires completed by parents. Those

questionnaires yielded significant pre- to post-treatment gains on measures of social competence. Tse and colleagues concluded that parent report suggested generalization to settings outside of the group, yet this was never evaluated directly by researchers. One potential benefit to social skills groups is the opportunity to practice skills with peers; however, other studies using social skills groups also reported poor maintenance and generalization of skills or did not directly assess for generalization and maintenance at all (Dotson et al., 2010; Nuernberger et al., 2013). The use of the social skills group format limits the opportunity for individualized instruction, which may account for the noted issues with generalization and maintenance.

Another common procedure used to teach social skills is behavioral skills training (BST). Behavioral skills training typically involves briefly explaining to the individual the rationale of the skill, modeling examples and nonexamples of the skill, then role play with feedback (Parsons et al., 2013). This approach may be more conducive to the acquisition of conversation skills because it involves individualized instruction with multiple opportunities for the participant to practice the skills with corrective feedback. Nuernberger et al. (2013) taught three young adults a sequence of greeting and conversation skills using BST. All three adults acquired the targeted skills; however, the conversation skills were taught in the context of a fixed conversation, so the sequence of the behavior was always the same. Nuernberger and colleagues (2013) did not assess for generalization of taught skills.

Chezan et al. (2020) also used a treatment package of BST, with the addition of in-vivo training, to teach conversation skills to three adults with ASD and moderate intellectual disability. Chezan and colleagues evaluated the impact of their treatment

package using a concurrent multiple-baseline design across participants. Their primary dependent variables were the percentage of steps of a conversation task analysis completed independently and the number of independent self-initiations of interaction with a peer. The behavioral skills training portion of the treatment package consisted of providing a rationale for the skill, modeling the skill, then role-play and feedback. In-vivo training was done with the participant and a peer (a coworker) using covert audio coaching (CVC). The experimenter was on an audio call with the participant, who had a discrete earpiece on so they could hear the experimenter. After an opportunity to engage in one of the target conversation skills was presented, the experimenter allowed 2-3 seconds for the participant to emit a correct response. Following an error or no response, the experimenter gave feedback via the audio call to the participant with further instructions on what to say next. This method of coaching allowed for the experimenter to discretely give immediate feedback without disrupting the flow of the conversation. Results indicated that all participants acquired the targeted conversation skills, self-initiations increased, and this progress was maintained at follow-up probes. The authors did not evaluate generalization to peers or other conversation partners outside of the participant's coworkers.

Overall, evidence of generalization is lacking across social skills research. One noteworthy study that attempted to address the issue of generalization of social skills was done by Hood and colleagues (2017). They also used BST followed by in-vivo training as part of a treatment package to teach two teenagers and one child with high functioning autism. The effects of the intervention were evaluated using a concurrent multiple baseline design across groups of responses (i.e., greeting skills and conversation skills)

for each participant. The treatment package initially included BST, trial-based teaching, role play, and reinforcement strategies. Behavioral skills training consisted of explaining the skill, the context in which it should be used, a rationale for why it's important, and modeling correct and incorrect responses. Participants were required to tact if the modeled response was correct or incorrect. If they responded correctly, the experimenter provided praise. If they were incorrect, the experimenter provided the correct response and explained why it was correct. The participant then practiced the skill in a discrete trial teaching format. Differential reinforcement of correct responses was used with two of the three participants.

The skills were then taught specifically in the context of a conversation. This in-vivo training condition was almost identical to the assessment condition (described above). The main differences between the conversation assessment condition and the training condition were the length of the sessions, the confederates used, and the use of prompts as errors occurred. The training conditions were shortened to ten minutes for two participants and five minutes for one participant. When the participant made an error, a textual prompt with specific feedback was made visible. The textual prompts were used as to not interrupt the natural flow of conversation. For two participants, differential reinforcement of correct responses through a token economy was used. Tokens were deposited in a clear bowl following an independent correct response. The use of a token economy allowed feedback to be given when a correct response was emitted without interrupting the flow of the conversation. Tokens were exchanged for preferred items or activities at the end of the conversation.

Results showed that multiple conversation skills generalized across settings and conversation partners for each participant in the study. Hood and colleagues argue that in-vivo better emulates a typical conversation than training each skill in a discrete trial teaching format such as BST. Thus, teaching social skills in this way may better facilitate the acquisition and generalization of contextual responding since individuals are required to discriminate which skill to use in the moment. However, Hood and colleagues did not evaluate the impact of the initial treatment package alone prior to implementing the in-vivo training component. Therefore, the extent to which in-vivo training enhanced generalization compared to the discrete-trial component of BST is unknown.

The Current Study

This study was an extension and replication of Hood and colleagues' (2017) method of assessing and treating social skills. The purpose was to further evaluate the impact of in-vivo training on generalization of conversation skills by administering generalization probes throughout the course of treatment. The modified assessment and intervention protocol aimed to address some of the potential limitations of previous direct assessments and interventions of social skills. Two adolescent girls with autism were recruited to participate in the study. Instead of a multiple baseline design, a multiple probe design across behaviors was used (Horner & Baer, 1978) to evaluate the impact of the intervention on the generalization of the skill across conversation partners. The initial assessment was probed throughout treatment as components were added. Each participant first experienced the modified assessment protocol for an initial probe. Then, three conversation skills were identified collaboratively using assessment results and input from the caregiver and participant. The chosen conversation skills were then targeted

using a treatment package of behavioral skills training, discrete trial teaching, and differential reinforcement. An assessment condition was probed throughout the training to assess the acquisition and generalization of trained skills. If the treatment package did not result in the acquisition and generalization of conversation skills, the skills were then taught in the context of a conversation with prompting and feedback (in-vivo training). The primary research question for this study was as follows: What is the relative impact of in-vivo training following BST on the generalization of conversation skills to novel conversation partners?

Method

Participants, Setting, and Materials

Two participants with autism spectrum disorder (ASD) were recruited to participate in the study. Participants were 14- and 13-years old girls with a diagnosis of ASD and deficits in social skills. Participants were recruited by distributing flyers to local Applied Behavior Analysis (ABA) providers and organizations in the area, and by posting flyers on social media pages for parents of children with ASD. To participate in the study, potential participants were screened according to the following inclusion criteria: 1) speaks in full sentences, 2) demonstrates the ability to engage in back-and-forth conversation when prompted, and 3) demonstrates the ability to remain engaged in a work task for up to 20 minutes. This was assessed via direct observation and interview using a checklist developed by the researchers (see appendix D).

Assessment and treatment sessions with Participant 1 took place in her home to accommodate transportation barriers cited by the participant and her family that would have otherwise impacted her ability to participate in the study. All sessions with

Participant 2 took place in a private session room at a university-based clinic. For both participants and settings, the session area included a table and two chairs, along with video recording equipment (a camera and tripod). The recordings of sessions were later watched by trained observers for data collection purposes.

Sessions were 2-3 hours long and occurred 1-2 days per week. Prior to each 20-minute (assessment) or 10-minute (training) conversation or set of trials, the researcher obtained assent from the participant by asking if they would like to start. The researcher then waited for an explicit statement from the participant indicating assent to begin the assessment or set of trials (e.g., “Yes”, “I’m ready to begin”, etc.). If assent was verbally withdrawn at any time (e.g., “I want to stop” or a similar statement was made by the participant), the researcher immediately discontinued implementing the assessment or treatment procedure until assent was re-given. If the participant displayed signs of dissent but did not explicitly state they were withdrawing assent (e.g., statements about wanting to leave or being bored, mild nonverbal indicators of irritation or boredom, etc.), the researcher reminded the participant that they could withdraw at any time and reaffirmed assent before continuing the session.

At the end of each assessment, conversation, or set of trials, the participant was given a 5–10-minute break to engage in a preferred activity of their choice. Access to these preferred items was not contingent on their level of performance during the session, nor were any contrived reinforcement contingencies (e.g., token economy) utilized during training sessions. Descriptive data was collected throughout the study on each participant’s affect and any signs of dissent observed during sessions. Participant 2

verbally withdrew assent to participate in the study during the in-vivo training phase with the primary assessment confederate. Sessions were immediately terminated in response.

The experimenters and several research assistants assisted with the implementation of the study, each serving in one of multiple roles. *Confederates* were the participant's conversation partners during probes and treatment phases. To assess generalization across multiple conversation partners, participants conversed with three different confederates in the probe condition, and a fourth confederate was only associated with training. Confederates associated with training were also responsible for providing textual prompts during in-vivo training (see treatment method section below for more details). *Observers* were trained data collectors who reviewed recorded sessions and collected data on the primary dependent variables (see section below for more details).

Dependent Variables and Interobserver Agreement

The primary dependent variables in the study were participants' performance on a semi-structured conversation assessments administered during baseline and as each participant met mastery criteria during training phases. Three conversation skills per participant were then selected to be targeted during BST and in-vivo teaching phases.

Data Collection

All conversations were video recorded. Observers collected data from watching the video on a number of different targets. Data was collected on the number of social presses delivered, as well as the number of correct responses to the social presses. The percentage of appropriate responses for each social press was then calculated by dividing the number correct by the total number of presses, then multiplying that number by 100.

The specific social presses assessed and treated varied for each participant and were selected based on direct observation and parent interview. See the table below for all the social presses that were targeted during the study. Data on additional targets were collected using both continuous and discontinuous measures (see Table 1 below for a list of targets and operational definitions).

<i>Participant 1</i>		
Data Collection	Target	Operational Definition
Percentage Correct (response to social press)	Responds to nonverbal social cue of boredom	The confederate uses body language or other overt behaviors (such as a change in facial expression, pitch/tone of voice, etc.) to indicate boredom Correct response: client vocally attends or responds to the nonverbal social cue in a way that is contextually and socially appropriate (by changing the conversation topic, asking what they would like to talk about, etc.) Incorrect response: client does not attend to the nonverbal social cue and continues with the conversation as normal or does not respond at all within ten seconds (e.g., confederate begins yawning and looking at their phone, participant continues conversation)
Frequency	Repeating past conversation	Asking a question that was just answered/discussed within the current 20-minute sample (e.g., Confederate: "It rains every other week in Jersey" Participant: "Does it rain a lot in Jersey?"), or repeating a statement either word for word or with the same content that was explicitly stated previously (e.g., tells confederate what her favorite color is more than once, mention of teacher's name more than once) within the current 20-minute sample. Does not include asking questions for clarification.
Frequency	Asking questions	A sentence worded or expressed as to elicit more information, usually (but not always) starting with who, what, when, where, why, how.
<i>Participant 2</i>		
Data Collection	Target	Operational Definition
Partial Interval (5 seconds)	Talking about self	The participant issues a statement about themselves (e.g., their interests, their past experiences, wants/needs, their friends and family, etc.) This usually includes "I, me, my" statements. Example: "I don't like the snow" Nonexample: "I heard tomorrow it is supposed to snow" (main content of that sentence is snow)
Frequency	Off-topic statements	Participant responds to confederate's statement or question with a statement or question with all content unrelated to the referent/topic mentioned by the confederate Nonexample: participant includes transition to introduce new topic (e.g., Confederate asks her favorite movie, she responds with "I don't watch movies, but my favorite music artist is...")
Percentage Correct (response to social press)	Responds to nonverbal social cue of boredom/annoyed	The confederate uses body language or other overt behaviors (such as a change in facial expression, pitch/tone of voice, etc.) to indicate feeling bored or annoyed Correct response: client vocally attends or responds to the nonverbal social cue in a way that is contextually and socially appropriate (by changing the conversation topic, asking what they would like to talk about, etc.) Incorrect response: client does not attend to the nonverbal social cue and continues with the conversation as normal or does not respond at all within ten seconds (e.g., confederate begins yawning and looking at their phone, participant continues conversation)

Table 1. Operational definitions and measures for selected target behaviors for each participant

Interobserver Agreement

Interobserver agreement (IOA) was collected using a second data collector for at least 30% of sessions for each participant (32% of Participant 1's sessions, and 37% of

Participant 2's sessions). Observers sampled an equal amount of sessions in each condition to ensure interobserver agreement across all treatment and assessment conditions. To calculate observer agreement on the social presses, the recorded responses of each data collector were compared to determine if the secondary data collector recorded the same target response occurring within 3 seconds before or after the primary data collector (MacLean et al. 1984). If both data collectors recorded the same behavior to have occurred within 3 seconds of the primary data collector, and both recorded it as a correct or incorrect response, it was endorsed as an agreement. The number of agreements was then divided by the number of agreements plus the number of disagreements, then multiplied by 100 to get a percentage of agreement. See Tables 2 and 3 below for the mean percentage and range of agreement scores for each dependent measure and condition for each participant.

<i>Participant 1</i>		
Dependent Measure	Interobserver Agreement Baseline	Interobserver Agreement Post-BST
Responding to nonverbal cues of boredom	93% (range, 89-95%)	91% (range, 88-94%)
Repeating past conversation	89% (range, 87-92%)	94% (range, 90-96%)
Asking questions	96% (range, 91-98%)	95% (range, 92-97%)
Affect	98% (range, 95-99%)	92% (range, 88-95%)
Indices of boredom and displeasure	100%	100%

Table 2. Interobserver agreement data (range and mean) for Participant 1 by dependent measure and condition

<i>Participant 2</i>				
Dependent Measure	Interobserver Agreement Baseline	Interobserver Agreement Post-BST	Interobserver Agreement Post-In-Vivo	Interobserver Agreement In-Vivo Training
Talking about self	89% (range, 86-93%)	83% (range, 79-91%)	85% (range, 81-90%)	87% (range, 83-96%)
Off-topic statements	95% (range, 92-98%)	93% (range, 91-96%)	91% (range, 89-92%)	97% (range, 92-98%)
Responds to nonverbal social cue of boredom/annoyed	93% (range, 88-94%)	94% (range, 90-96%)	89% (range, 87-92%)	90% (range, 86-91%)
Affect	89% (range, 84-93%)	86% (range, 81-88%)	93% (range, 88-94%)	89% (range, 87-92%)
Indices of boredom and displeasure	87% (range, 82-89%)	77% (range, 73-81%)	85% (range, 77-87%)	84% (range, 78-86%)

Table 3. Interobserver agreement data (range and mean) for Participant 2 by dependent measure and condition.

Experimental Design and Procedures

Following the initial assessment, the effects of BST and in-vivo teaching on increasing and promoting generalization of conversation skills were evaluated using a multiple probe across behaviors design (Horner & Baer, 1978). Following baseline, participants experienced BST, followed by assessment probes, and then in-vivo teaching if needed. Probes were administered periodically throughout each phase, and the next phase was not introduced until stable performance was observed.

Initial Assessment

Eligibility to participate in the study, as well as conversation skills to target during assessment, were first identified through caregiver and participant interview, followed by direct observation. Caregiver interviews were conducted in a private room with only the researcher and caregiver present. The researcher asked the caregiver open-ended questions about what their child's strengths and weaknesses in conversations were, and what goals they thought were important for their child to work on. See appendix B for the questions the researcher asked during the caregiver interview.

Following the caregiver interview, the participant was asked to enter the room to begin the direct observation. The direct observation lasted approximately 30 minutes. First, the researcher gave the participant and caregiver a series of prompts (e.g., "discuss a topic your child enjoys talking about") and instructed participant and caregiver to interact and converse as they normally would. After about 15 minutes, the researcher then interacted with the participant while the caregiver stepped out of the room. The researcher asked the participant to identify what topics to discuss during the interaction. If a low rate of initiations or responding was observed, the researcher asked the participant general questions about themselves and introduced topics identified as preferred by the caregiver.

The researcher then interviewed the participant on what they felt were their strengths and weaknesses during conversations, and what was important to them to work on. See appendix C for the questions the researcher asked during the interview. Based off the direct observation and caregiver report, the researcher determined eligibility to participate in the study. See appendix D for the form used.

Baseline and Assessment Probes

A direct assessment protocol, modified based on the Hood et al (2017) study, was then administered to directly assess performance on the identified targets during baseline and throughout the course of the study. Participants experienced three 20-minute semi-structured conversation conditions with a confederate. Throughout the conversation, a variety of “social presses” were delivered by the confederate, which served the same function as the “evocative situations” embedded in the Hood et al (2017) study. For each participant, two of the social presses included a greeting and closing statement from the confederate at the beginning and end of the conversation. Given the complexity and wide range of other conversation skills to target, other social presses individualized to each participant were selected during the direct observation and parent/caregiver interview. Using social presses that were selected for each individual participant, as done in the Hood et al. (2017) study, allowed for a flexible assessment protocol that could better capture each individual’s skill deficits in comparison to a standard set of social presses.

Confederates were instructed to embed each social press for a predetermined minimum number of times during the conversation. Confederates were also instructed to allow the conversation to be participant-driven. If the participant did not introduce a topic or there was a pause in the conversation greater than 10 seconds, the confederate would introduce a topic. Confederates were given information about the participant’s preferences and preferred/nonpreferred topics prior to their first meeting. See appendix A for a sample reference sheet confederates used during the assessment to ensure all social presses were embedded. The confederate associated with training condition differed from confederates associated with the assessment condition to assess for generalization.

Certain social presses were more likely to be naturally embedded more than the minimum number of times during the conversation (e.g., asking a question). Since the greeting and goodbye could only be embedded into the conversation once, each assessment was delivered three times as opposed to once in order to get a more thorough sampling of the participants' current repertoires.

Following the initial baseline assessment, three conversation skills per participant were selected to move to the treatment phases. Targets were selected based on the caregiver and participant's reported primary concerns, and the participant's performance on the baseline assessment. The lowest percent correct scores obtained in response to the delivered social presses were first considered for treatment. Ultimately, the caregiver and participant aided in the final selection of targets.

Behavioral Skills Training (BST)

The initial intervention for each target behavior was behavioral skills training (BST). Each behavioral skills training session began with an explanation of the skill, a description of the context in which it should be applied, and a rationale as to why it's important in a conversation. During the first phase of BST, the training confederate modeled correct and incorrect examples of the skill and asked the participant to identify whether the model of the skill was correct. Once the participant demonstrated 90% or better performance with tacting the accuracy of the training confederate's model across three consecutive sessions, they were progressed to the next phase of BST.

During the final phase of BST, participants were asked to role-play the skill with the researcher. Upon an incorrect response, a prompt card was held up. Each prompt card was 8cm by 28cm in size and had a picture that was associated with specific feedback

(i.e., a picture of arrows was used as a prompt to the participant to change the topic, and a picture of a straight horizontal line with a period next to it was used as a prompt to issue a statement). We did not conduct any pre-training related to the prompt cards; rather, specific feedback (i.e, “Change the topic”, “Issue a statement”) was paired with each presentation of the prompt card each time the participant made an error. At the end of each role play, the confederate reviewed their score and any corrective feedback with the participant. After 90% or better performance across three consecutive sessions of the role-play phase was met, the full conversation assessment across at least three (or until stability is observed) 20-minute sessions was re-administered to evaluate whether the BST treatment package was enough to result in the generalization of those skills to a conversation with three confederates. If performance on the assessment probe was below criterion, the participant then experienced in-vivo teaching for the target skill.

In-vivo Teaching

During the in-vivo teaching phase, the participant engaged in a conversation with the training confederate to practice the targeted skill. The semi-structured conversations were delivered the same as described in the assessment procedures above, except the sessions were shorter in duration (10 minutes), and the picture prompts used during role-play were delivered to the participant following an incorrect response. To avoid interrupting the natural flow of the conversation, prompts were only delivered visually (e.g., in the absence of any verbal specific feedback from the researcher) using the same prompt cards as mentioned above. The card was held up by the researcher while they continued the conversation. The researcher did not provide differential consequences during the conversation following a correct response.

Once responding for a particular skill was demonstrated to mastery with the use of picture prompts, which were individualized for each participant, the participant was again exposed to the probe described above.

Procedural Integrity

At least 30% of sessions for each participant (33% of Participant 1's sessions and 37% of Participant 2's sessions) were observed for procedural integrity purposes. A relatively equal amount of training and probe sessions were observed to ensure treatment integrity across all phases. See Table 4 for the range and mean of procedural integrity scores obtained during each phase for each participant.

Each component of all phases was broken down into a checklist. Observers endorsed Y if the confederate implemented a particular step correctly for the entirety of the session and endorsed N if at least one error was made. Total treatment integrity for the session was calculated by dividing the number of tasks endorsed with Y by the total possible and multiplying that number by 100 to get a percentage.

During the probe and in-vivo training condition, observers tracked whether the confederate gave the participant 10 seconds to initiate a greeting, embedded all target social presses, ended the conversation after 10 minutes, provided any feedback to the participant, and kept all statements general and uncomplicated. Observers also noted each social press that was embedded, as well as if the press was delivered correctly per the operational definition provided, if the confederate allowed up to 10 seconds for the participant to respond, and if the appropriate picture prompt card was delivered following an error (in-vivo training condition only). During BST sessions, observers collected data on whether the confederate ran the appropriate amount of trials, completed all steps of

BST accurately, and implemented prompting procedures correctly. See appendices E and F for samples of procedural integrity checklists used for each phase.

<i>Participant 1</i>	
Condition	Procedural Integrity
Baseline	88% (range, 86-92%)
BST Training: Responding to nonverbal cues of boredom	97% (range, 94-100%)
Post-BST Probe: Responding to nonverbal cues of boredom	98% (range, 96-99%)
BST Training: Asking questions	93% (89-97%)
Post-BST Probe: Asking questions	98% (range, 96-100%)
<i>Participant 2</i>	
Condition	Procedural Integrity
Baseline	89% (range, 87-90%)
BST Training: Responding to nonverbal cues of boredom/annoyed	96% (range, 90-98%)
Post BST Probe: Responding to nonverbal cues of boredom/annoyed	98% (range, 93-100%)
In-vivo Training (training confederate): Responding to nonverbal cues of boredom/annoyed	91% (range, 88-93%)
Post-In vivo Probe: Responding to nonverbal cues of boredom/annoyed	92% (range 89-93%)
In-vivo Training (confederate 1): Responding to nonverbal cues of boredom/annoyed	92% (range, 87-94%)

Table 4. Mean procedural integrity scores by participant and condition.

Social Validity

The caregiver and participant were asked to complete social validity forms (see appendices H, I, J, and K) following the initial assessment as well as at the end of the study. The social validity forms asked the participant and caregiver to rate how much they agreed with statements related to their level of satisfaction with the procedures, goals, and outcomes of the study on a Likert-type rating scale (Parker et al., 2022; Wolf, 1978). The scale on the caregiver form ranged from 0-6 (0 being strongly disagree and 6 being strongly agree). The scale on the participant social validity form had 3 options: “No” (1), “A little” (2), and “Yes” (3) with an image of a frown face, neutral face, and smiling face under each option, respectively. At the end of each form there was a section that allowed the participant and caregiver to add any additional comments they had related to the study,

At the end of the study, the participant and their caregiver were asked to watch a video of the direct assessment administered during the baseline phase, as well as a video of the assessment administered at the end of study. A section of the post-study social validity survey asked the participant and caregiver to rate their agreement to statements related to the video and the extent to which they observed a noticeable improvement of the participant’s performance on the targeted conversation skills.

Affect Data

All affect data were collected as part of a post-hoc analysis of the general affect of each participant to aid in a descriptive analysis of potential indicators of assent and withdrawal of assent for both participants. As such, these data are not complete.

Following baseline, descriptive data were collected by the researcher during each session on Participant 2's general affect, whether she made statements indicating she was bored, and whether there were overt indicators of displeasure. Specifically, Participant 2's general affect was noted as either positive, negative, or neutral by the researcher at the end of each session. A positive affect was defined as the participant displaying overt indicators of pleasure (e.g., smiling, laughing, explicit statements of happiness or pleasure, etc.) for the majority of the session. A negative affect was defined as the participant displaying overt indicators of displeasure (e.g., rolling her eyes, sighing, explicit statements indicating boredom or displeasure, etc.) for the majority of the session. A neutral affect was defined as an overall absence of overt indicators of both pleasure and/or displeasure for the majority of the session. Statements of boredom were defined as at least one statement made to the researcher or research assistants that explicitly indicated the participant was bored or disinterested (e.g., "This is boring", "I don't like doing this", etc.). Finally, indices of displeasure were defined as one or more overt indicators of displeasure (e.g., rolling her eyes, sighing, etc.) occurring during the session.

These data were collected in response to Participant 2 withdrawing assent during the last baseline session, as well as general observations of Participant 2 having a negative affect during sessions at the start of the study. The purpose of the data was to evaluate and document potential overt indicators of assent withdrawal to aid in decision making with whether to continue with the study with Participant 2. With the exception of one explicit statement indicating she would like to be finished early for the day during the last initial baseline session, Participant 2 consistently gave explicit verbal assent when

asked, despite frequent occurrences of negative affect and signs of displeasure. Assent was consistently reaffirmed immediately following an indicator of displeasure, per the assent procedures described above.

A post-hoc analysis of Participant 1's affect and indices of boredom or displeasure was conducted following completion of the study to verify the validity of the affect data collected for Participant 2; given Participant 1's consistent positive affect anecdotally observed across all researchers, we expected that the same data for Participant 1 would look quite different. Participant 1's data was collected by watching all sessions that were recorded (each probe session was recorded for data collection purposes, and at least 30% of training sessions were recorded for procedural integrity purposes). The recordings used to collect data for Participant 1 only depicted when the participant was working with the assessment confederates or training confederates (i.e., only during work trials or assessment periods, not during noncontingent break periods or session set up), while Participant 2's data represents her affect and indicators of boredom/displeasure throughout the entire session, including during noncontingent break periods. See the results section below for a visual display of the data collected for both participants.

Results and Discussion

Results

See Figures 1 and 5 below for a graph of results for the study for participants 1 and 2, respectively. The y-axis depicts each participant's performance for each target behavior during probes administered throughout each treatment phase. The x-axis depicts sessions. The open circle, triangle, and square data paths represent the participants'

performance with confederates 1, 2, and 3 respectively, all of which were only associated with the probe condition (separate from the training confederate).

Figures 2, 3, (Participant 1) and 6 (Participant 2) depict data obtained during training sessions for BST and in-vivo teaching phases. The x-axis represents sessions, while the y-axis represents the participant's average percent correct score for all trial blocks ran for that session. The asterisks on Participant 2's charts (Figures 4 and 5) indicate when she began attending a social skills group 3 days per week during the study.

Figures 4 (Participant 1) and 7 (Participant 2) depict each participant's affect and whether there were any indicators of boredom or displeasure during each session on a cumulative graph. The x-axis represents each day in which a session occurred, while the y-axis depicts a cumulative number of each affect category (note that the affect categories were mutually exclusive, each affect could only be scored once per session). Each type of affect is represented on its own data path; the closed triangle data path represents sessions in which participants showed an overall positive affect, the solid squares negative affect, and the closed circles neutral affect. The data paths are separated by phase change lines to indicate when each participant entered a new phase of the study. The breaks in data path during the BST: Nonverbal Cues of Boredom and BST: Modeling phases on Figure 4 (Participant 1) indicate sessions in which affect data were not collected for Participant 1 because the session was not recorded.

Figure 8 (Participant 2) depicts any session in which Participant 2 emitted at least one statement of boredom (the data path with the closed circles) and/or indices of displeasure (the closed squares data path). Statements of boredom and indices of displeasure were not mutually exclusive and could both be scored within the same

session. The x-axis represents each day in which a session occurred, while the y-axis depicts a cumulative number of statements of boredom and indices of displeasure. The data paths are separated by phase change lines to indicate the different conditions of the study that Participant 2 experienced. This data for Participant 1 is not depicted below because no statements of boredom or indices of displeasure were recorded as having occurred during any of the footage of sessions that was observed.

Participant 1

During baseline, Participant 1 responded at 0% accuracy to nonverbal cues of boredom issued by all 3 confederates. Following BST, Participant 1's accuracy increased to 100% and remained at 100% across confederates for the remainder of the study. In-vivo teaching was not introduced for this target, given that her behavior generalized to the novel conversation partners following BST. For asking questions, Participant 1 asked 44-84 questions during each initial 20-minute baseline sample. This target behavior decreased in level (ranging from 16-44) while still in baseline during the next sample. Following BST, asking questions further decreased, ranging from 12-25 questions per sample across all confederates. As with the first target behavior, in-vivo teaching was not introduced, as the change in the participant's behavior generalized to the conversation partners following BST alone. The third target behavior, repeating past conversations, decreased to desired levels while in the baseline condition; no BST or in-vivo teaching was introduced for this target.

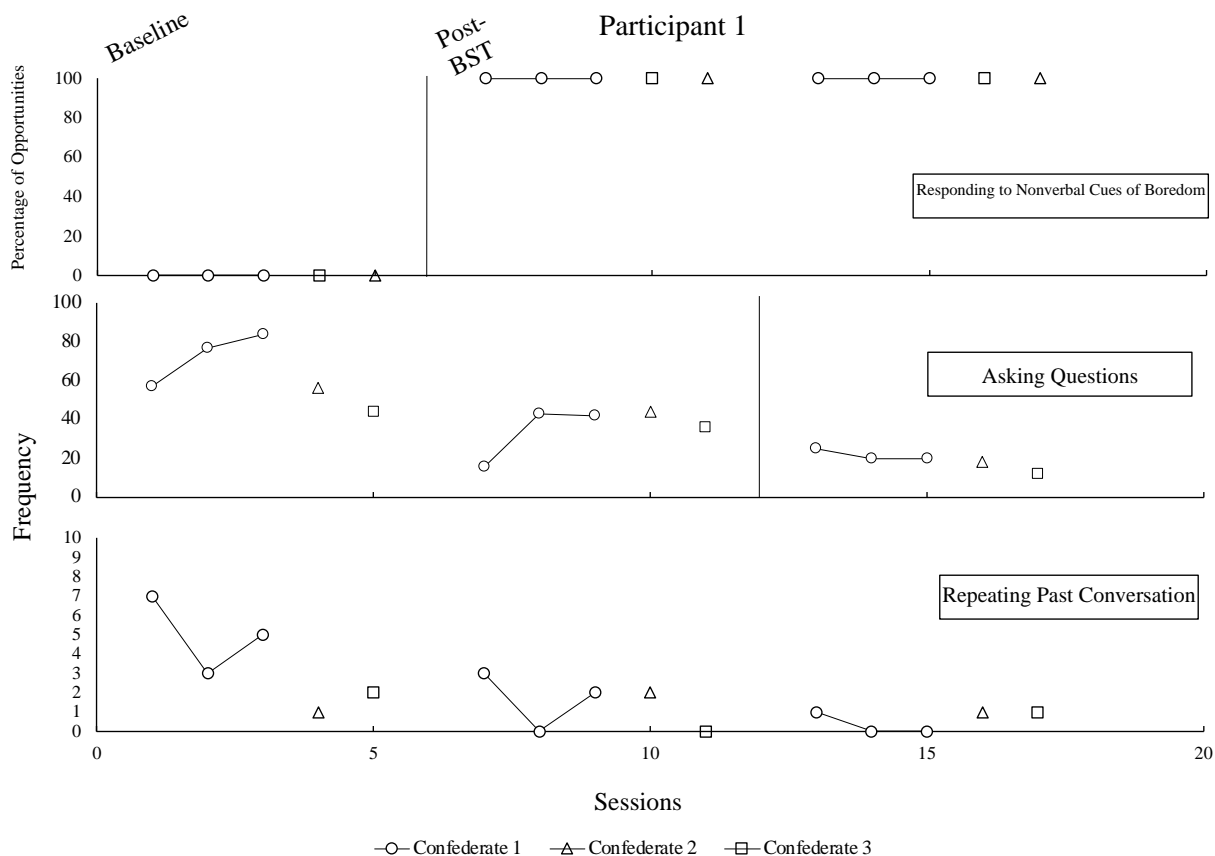


Figure 1. Participant 1's performance on the direct assessment probes throughout each phase of the study.

During BST training for responding to nonverbal cues of boredom with Participant 1, the participant met criteria for each phase of the skill in 1-2 sessions per phase. Only one session of the first phase was run. During the ID boredom phase, the participant was asked to identify whether the training confederate was displaying overt indicators of boredom. During the baseline probes (Figure 1) for this skill, the participant was observed tacting that the confederate was bored while the social press was issued. While she correctly identified that the confederate was displaying indicators of boredom, she did not change the topic, therefore she did not meet criteria for a correct response per the operational definition. Given her accuracy in identifying that the conversation partner

was bored during baseline and the high accuracy (100%) shown during the first session, the modeling phase was introduced following only one session of the initial phase. During modeling and role-play phases, the participant met criteria following only two sessions.

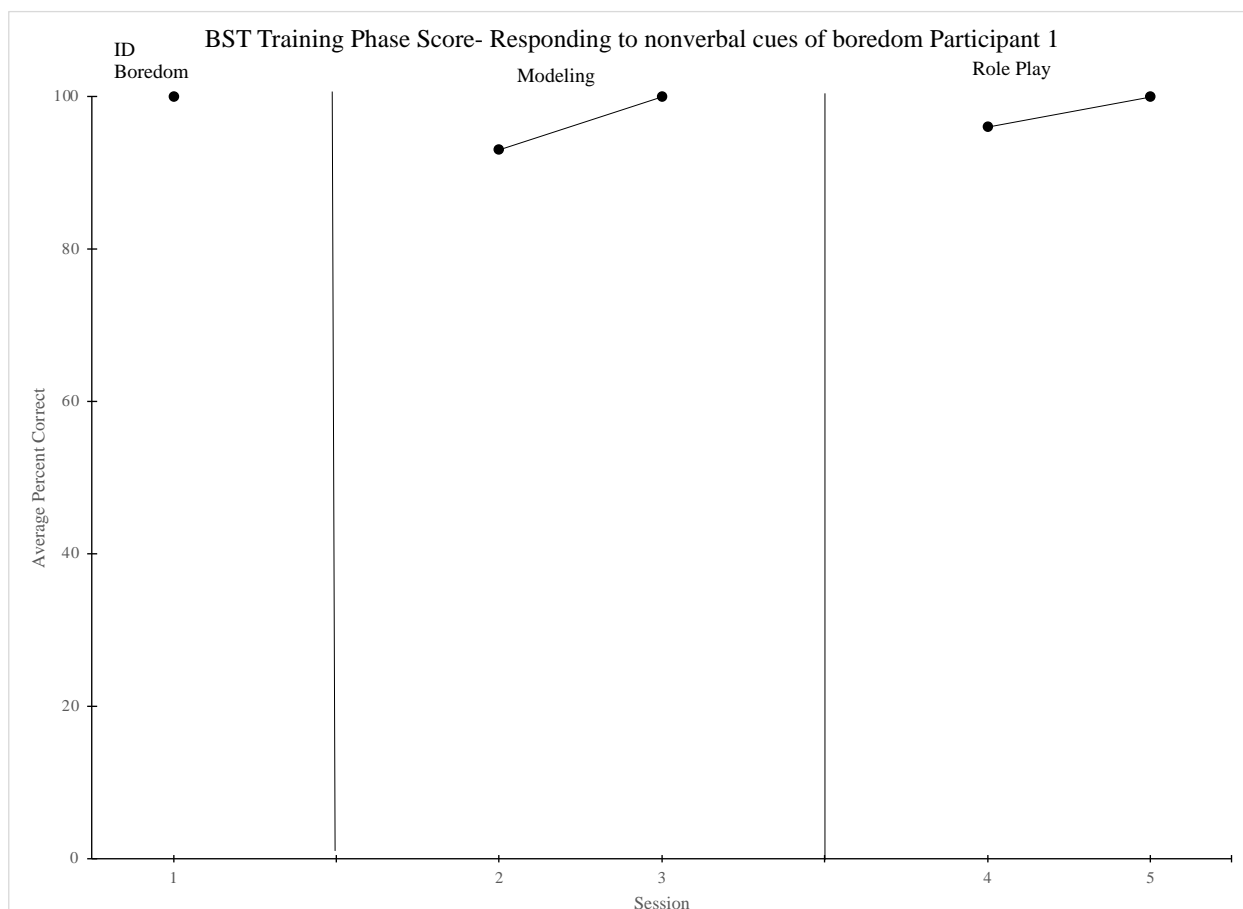


Figure 2. Participant 1's performance during each phase of BST training for the responding to nonverbal cues of boredom target.

Participant 1 was also quick to meet mastery criteria during BST for reducing question asking. During the modeling phase, the participant responded at 100% accuracy across the first two sessions. During the role play phase, the participant also met criteria following the first two sessions.

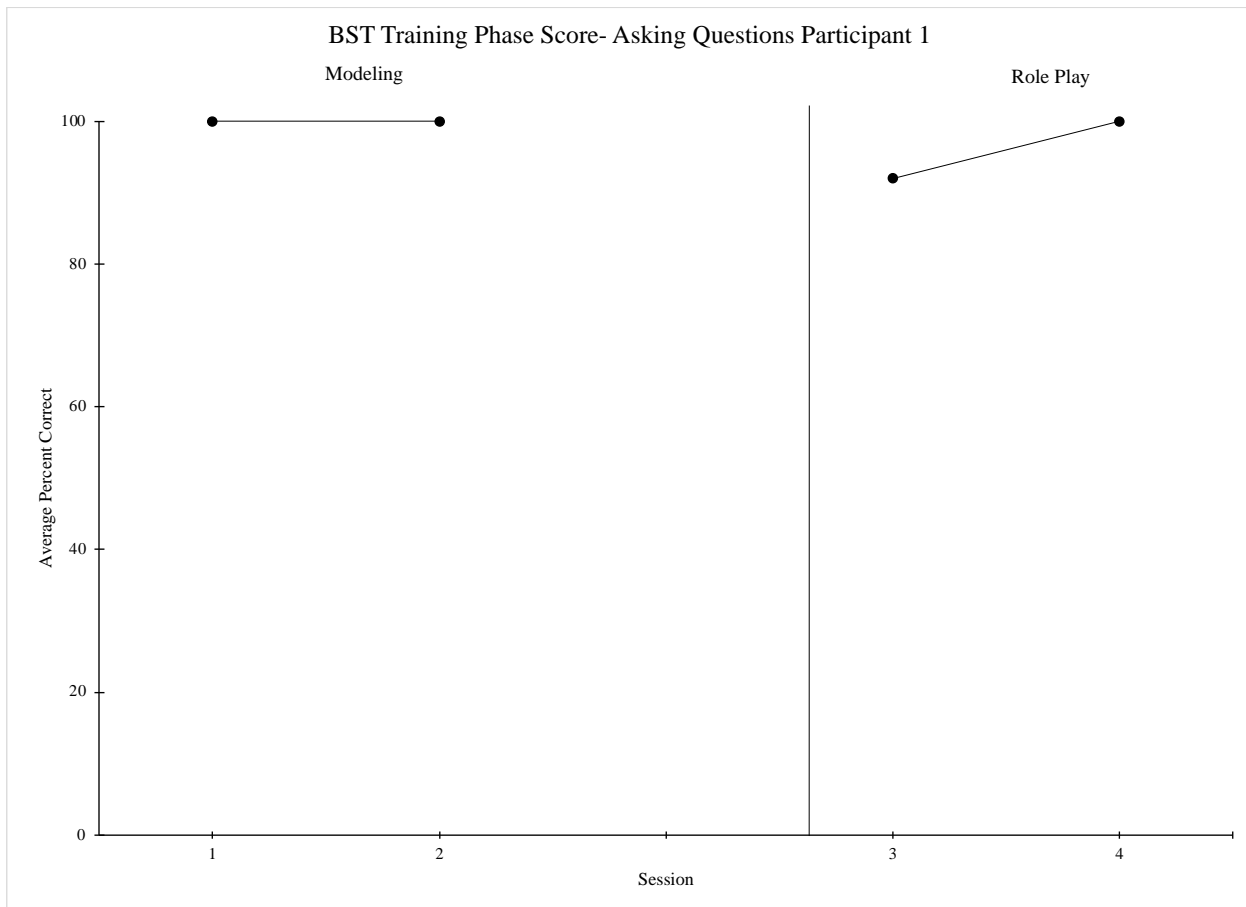


Figure 3. Participant 1's performance during each phase of BST training for the asking questions target.

Based on the post-hoc analysis of Participant 1's affect throughout the study, Participant 1 consistently showed a positive or neutral affect during each session across phases. The phase of the study with the most sessions in which Participant 1 displayed a neutral affect was during the BST: Nonverbal Cues of Boredom phase. During prior and subsequent phases, a steep increase in sessions in which Participant 1 displayed a positive affect was observed.

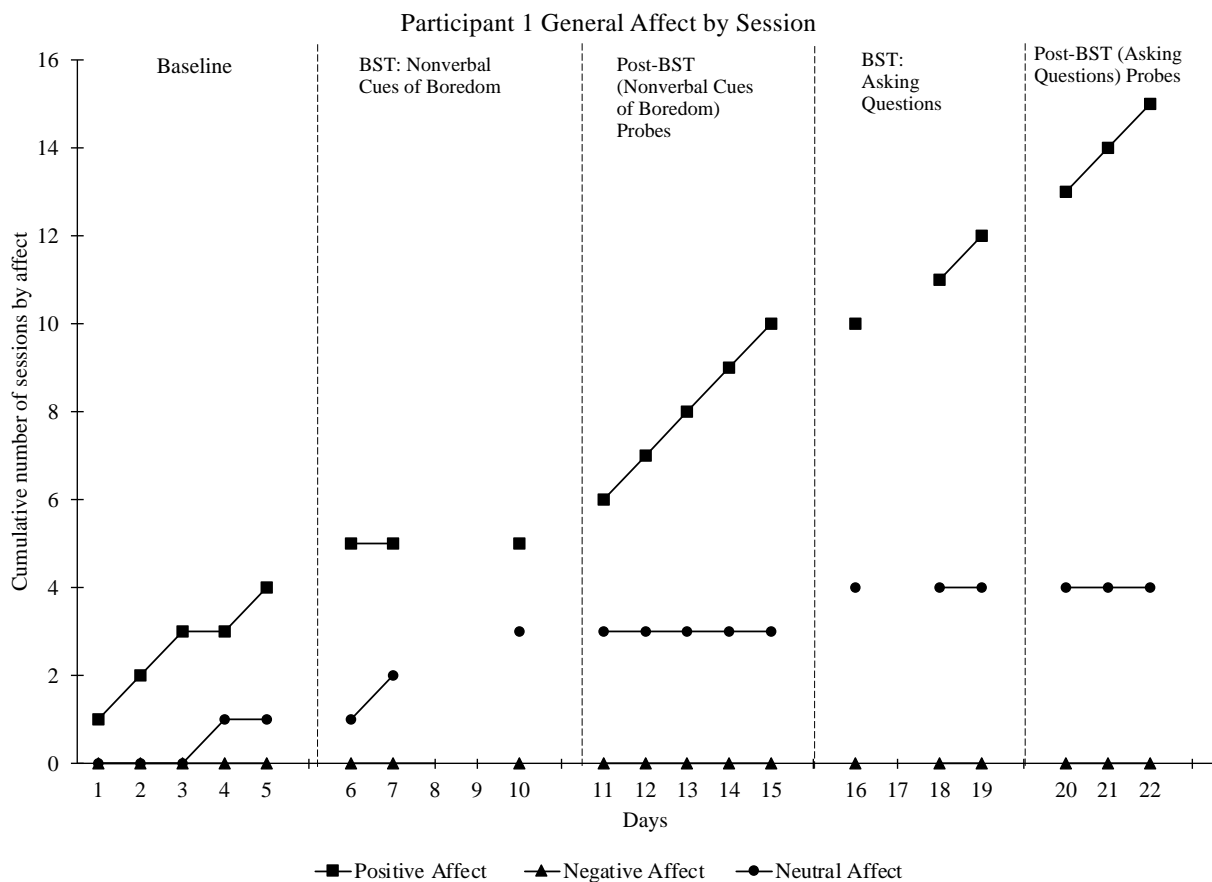


Figure 4. This graph depicts the cumulative number of sessions throughout the study in which Participant 1's general affect was scored as positive, negative, or neutral. The breaks in data path during the BST: Nonverbal Cues of Boredom and BST: Modeling phases indicate sessions in which affect data were not collected for Participant 1 because the session was not recorded.

Participant 2

Participant 2 showed low levels of accuracy during the baseline phase for responding to nonverbal cues of boredom/annoyed. Following BST training, accuracy in responding increased to about 33%, indicating that BST was not sufficient to produce robust generalization across three novel conversation partners. No change in level was observed following in-vivo training with the training confederate, which suggests that in-vivo training was also not sufficient to produce generalization of the skill. In-vivo

training was then introduced with the primary assessment confederate (Confederate 1), which improved responding. Before the skill reached criteria with Confederate 1, Participant 2 withdrew assent, and we ended the study with this participant. The other two target behaviors, off-topic responses and talking about self, remained in baseline throughout the course of the study. Off-topic responses showed a slight decrease in level throughout the course of the study, but still tended to occur around 2-4 times per 20-minute sample. The percentage of intervals in which the participant spent talking about herself remained high, with some variability observed following the introduction of in vivo teaching with Confederate 1 for the first target behavior.

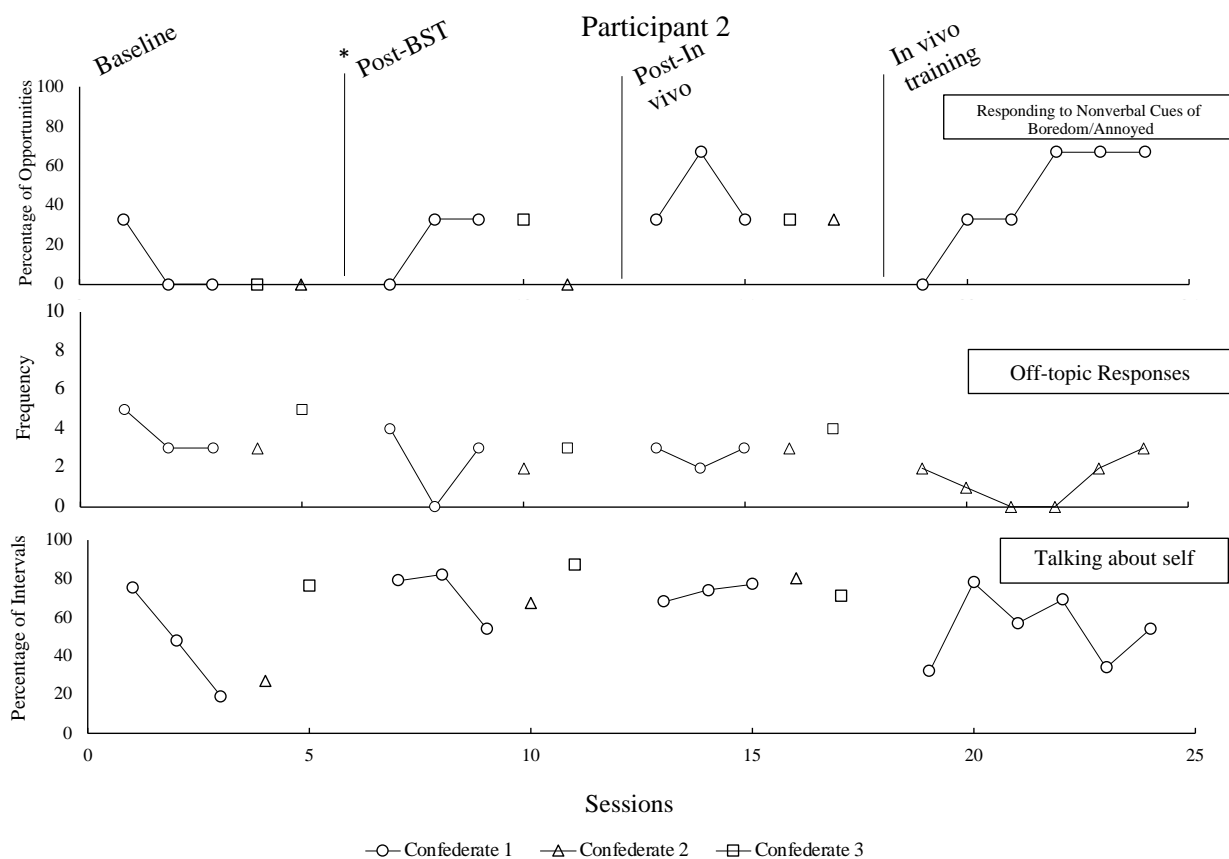


Figure 5. Participant 2's performance on the direct assessment probes throughout each phase of the study. The asterisk

on the graph indicates when Participant 2 began attending a social skills group three days a week with a local ABA provider.

Figure 5 below depicts Participant 2's performance during the training phase for the first target behavior. Across all BST phases, Participant 2's performance started at 65-82% accuracy and met mastery criteria after 4-5 sessions with a steady increasing trend. During in-vivo training, an initial drop in level is observed, and 10 sessions were required to train the skill to mastery.

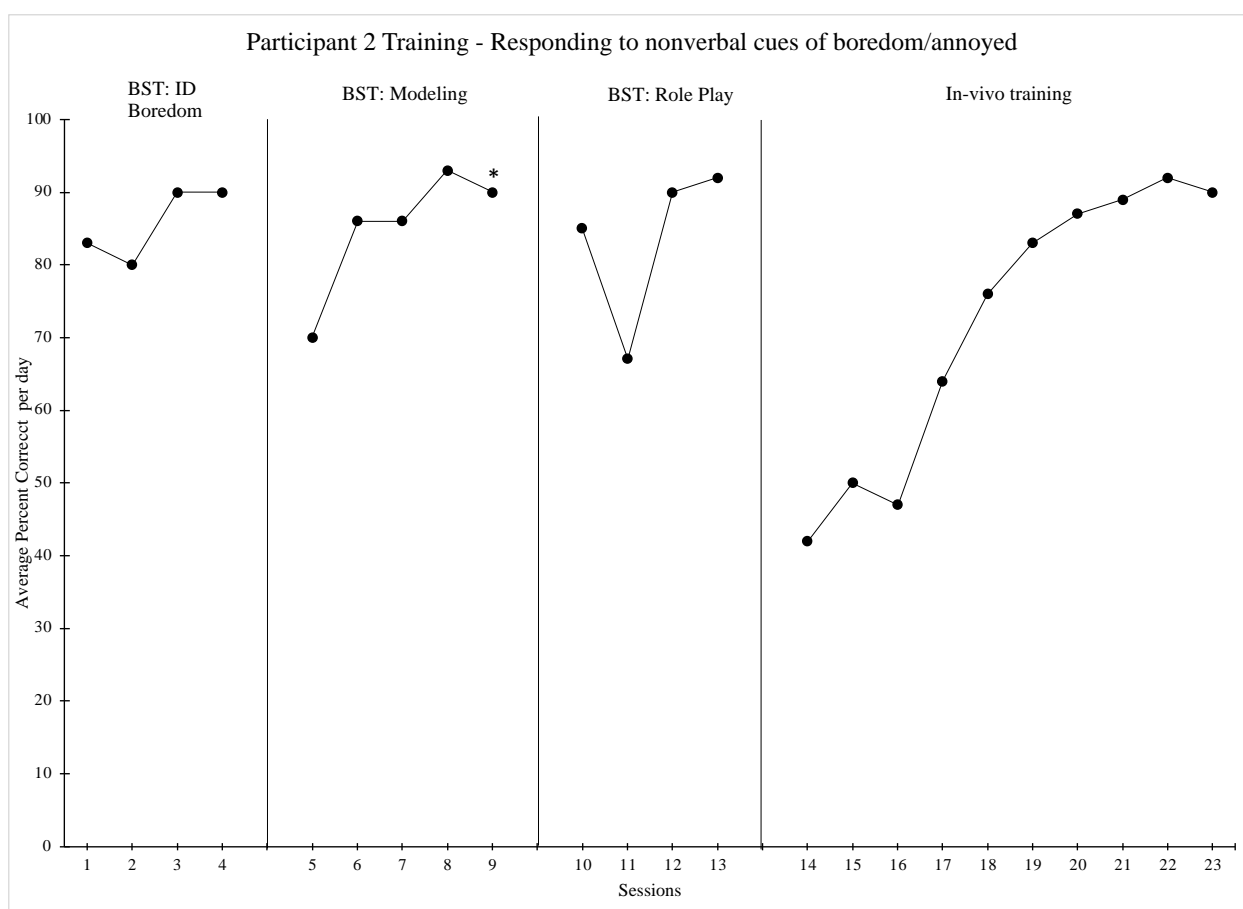


Figure 6. Participant 2's performance during each phase of BST training for the responding to nonverbal cues of boredom/annoyed target.

Figure 7 depicts the descriptive data obtained on Participant 2's general affect during each session. The descriptive data collected during each session was reviewed and

coded as either positive, negative, or neutral. For example, if the researcher noted that the participant was smiling and initiating social interaction with the researcher frequently throughout the session, that was considered a day in which the participant showed a positive affect. If for most of the session the participant displayed overt indicators of displeasure or boredom (e.g., statements indicating displeasure or boredom, rolling her eyes, sighing, etc.), that was considered a day in which the participant showed an overall negative affect. A session was coded as neutral if the participant did not display overt indicators of a positive or negative affect for most of the session.

Overall, Participant 2 showed a neutral affect for most of the sessions during the study. The only conditions in which the researcher noted sessions that Participant 2 had a positive affect was during the post-in vivo training probes and the in-vivo training with Confederate 1, which were the last two conditions that Participant 2 experienced. Interestingly, the condition with the greatest number of sessions where a negative affect was noted was also during the in-vivo training condition.

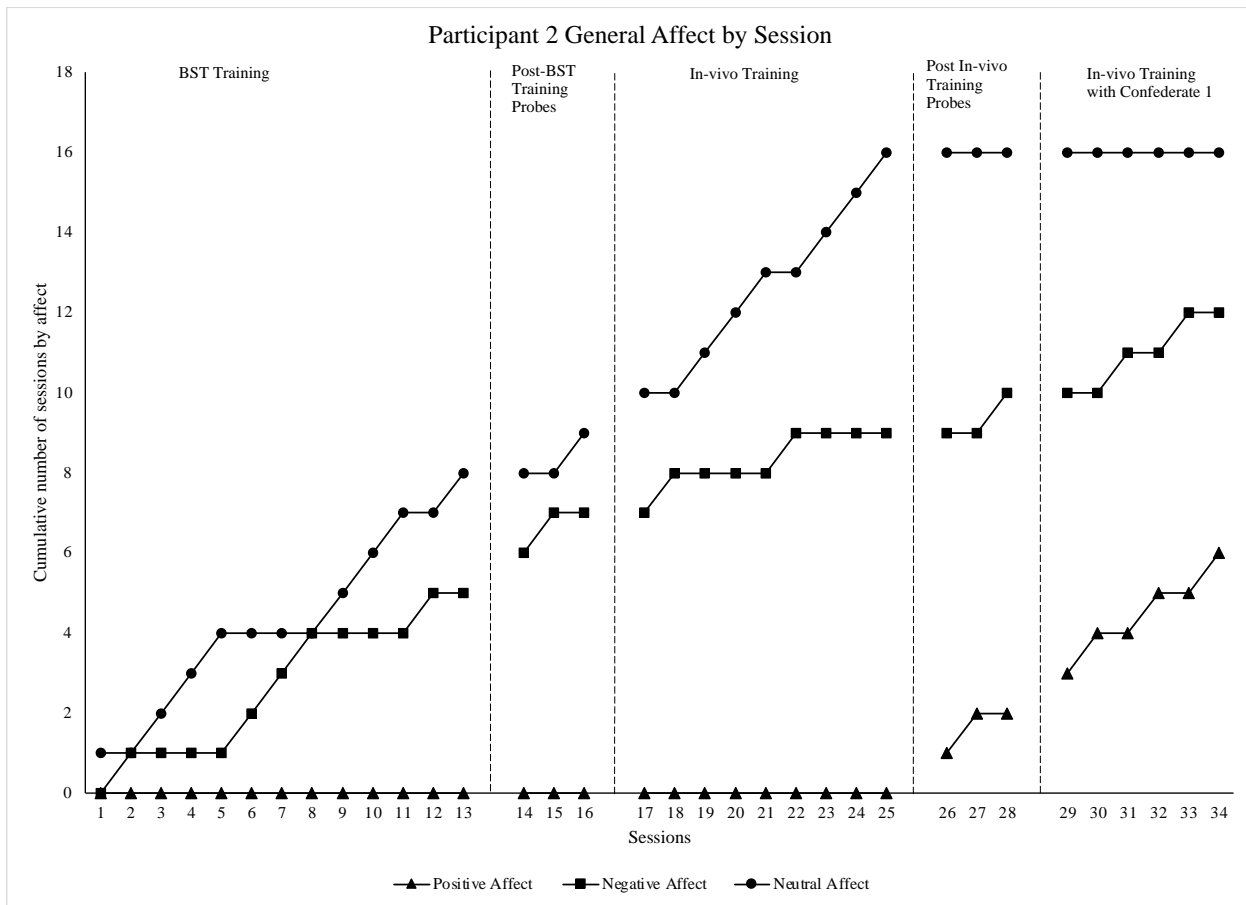


Figure 7. This graph depicts the cumulative number of sessions throughout the study in which Participant 2's general affect was scored as positive, negative, or neutral.

Figure 8 depicts an overview of the cumulative number of sessions in which Participant 2 made statements of boredom and/or displayed indices of displeasure. An occurrence of statements of boredom were noted if Participant 2 made one or more statements to the researcher and/or confederates that explicitly stated that she felt bored or disinterested while participating in the study. Indices of pleasure was defined as any session in which Participant 2 engaged in behaviors noted by the researcher that overtly indicated she may have been experiencing displeasure, such as making negative statements related to the social validity of the study (e.g., stating she did not enjoy

attending sessions with the researcher) or giving non-vocal cues of displeasure (rolling her eyes, sighing, etc.).

All conditions had at least one session where statements of boredom were noted by the researcher in the descriptive data collected. Statements of boredom were recorded as having occurred during all sessions of the in-vivo training condition with Confederate 1, the last condition Participant 2 participated in. During all other conditions, there was a relatively equal distribution of sessions with and without statements of boredom occurring. The condition with the highest number of sessions where displeasure was indicated by the participant was again the in-vivo training condition with Confederate 1, the last condition of the study for Participant 2.

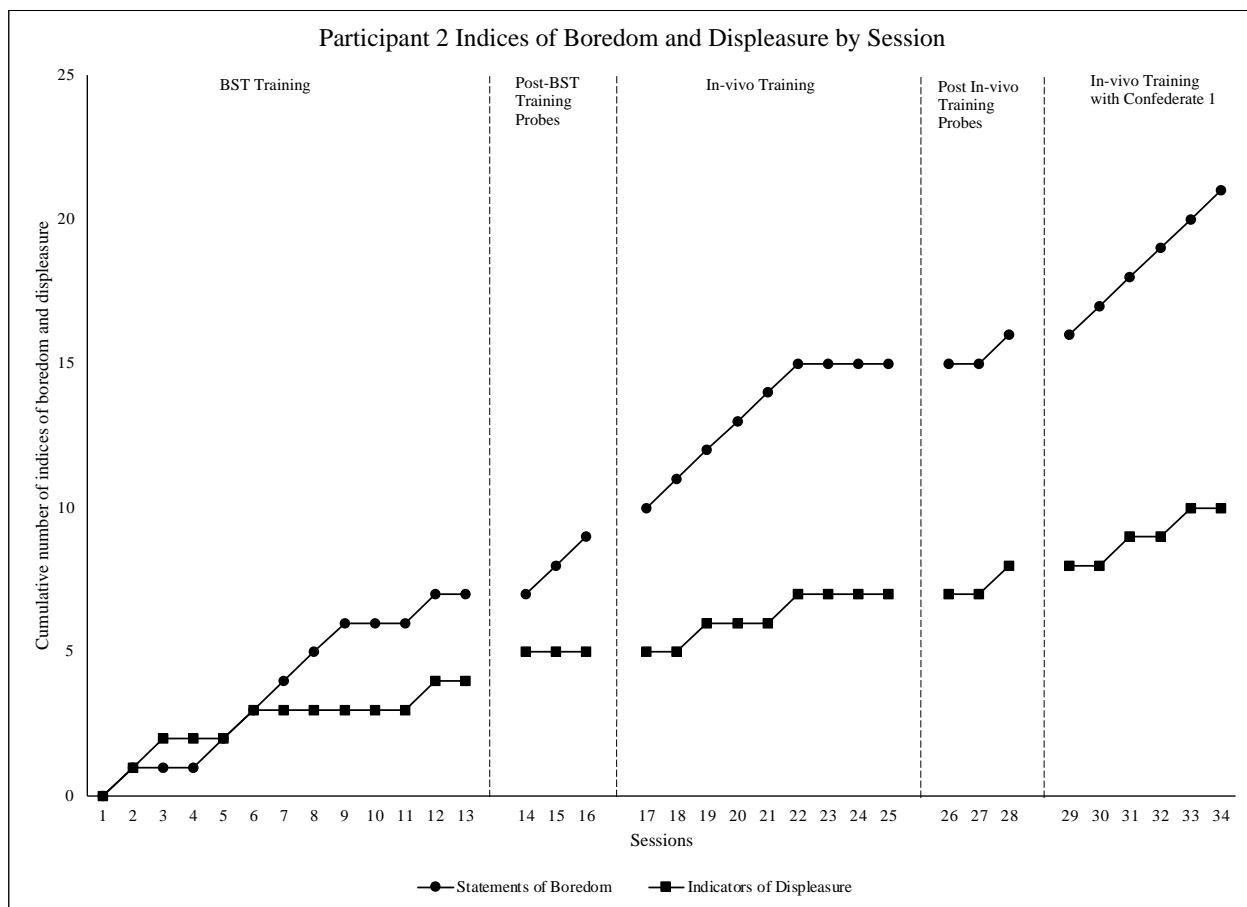


Figure 8. This graph depicts the cumulative number of sessions in which the researcher noted statements of boredom and/or indices of displeasure from Participant 2.

Tables 5 and 6 below show the questions asked and social validity ratings given in each category by Participant 1 and her caregiver. Participant 1 and her caregiver both reported high levels of satisfaction with the goals and intervention procedures post-assessment, as well as high levels of satisfaction with the effects of the intervention post-study. They both reported they observed a noticeable difference in performance following baseline to post-treatment.

<i>Participant 1 Pre-treatment Social Validity Rating</i>		<i>Participant 1 Caregiver Pre-treatment Social Validity Rating</i>	
<i>Goals</i>		<i>Goals</i>	
Question	Rating (“No”, “A little”, or “Yes”)	Question	Rating (scale of 0-6)
The researcher listened to me about what my goals for conversation skills were.	Yes	I feel included in the selection of goals for my child.	6
I feel that overall, the goals selected to work on are important to me.	Yes	I feel that overall, the goals selected to work on with my child are important.	6
It’s important to me to work on my conversation skills.	Yes	I feel that conversation skills are an important skillset for my child to work on.	6
I feel that responding to nonverbal cues of boredom is an important skill to work on.	Yes	I feel that responding to nonverbal cues of boredom is an important skill to work on.	6
I feel that staying on topic during a conversation is an important skill to work on.	Yes	I feel that using more statements during a conversation is an important skill to work on.	6
I feel that showing interest in others during a conversation is an important skill to work on.	Yes	I feel that avoiding repeating past conversations is an important skill to work on.	6
<i>Intervention</i>		<i>Intervention</i>	
Question	Rating (“No”, “A little”, or “Yes”)	Question	Rating (scale 0-6)
I understand how I will be taught the conversation skills.	Yes	The researcher explained the treatment procedures to me in a way that I understood.	6
How I will be taught conversation skills is appropriate based on my goals.	Yes	The methods that will be used to teach my child conversation skills are appropriate based on their goals.	6
How I will be taught conversation skills is appropriate for me and my needs.	Yes	The methods that will be used to teach my child conversation skills are appropriate based on my child’s needs.	6
<i>Comments</i>		<i>Comments</i>	
Question	Response	Question	Response
Please write anything else about things you liked and didn’t like about the initial assessment.	N/A; this section was left blank	Please write any additional comments regarding your satisfaction with the initial assessment	N/A; this section was left blank

Table 5. Questions asked and ratings given on the social validity form issued to Participant 1 and her caregiver pre-treatment.

<i>Participant 1 Post-Study Social Validity Rating</i>		<i>Participant 1 Caregiver Post-Study Social Validity Rating</i>	
<i>Effects of Intervention</i>		<i>Effects of Intervention</i>	
Question	Rating (“No”, “A little”, or “Yes”)	Question	Rating (scale of 0-6)
I enjoyed meeting with the researchers during the study.	Yes	My child enjoyed meeting with the researchers during the study.	6
I would participate in a similar study or intervention program again.	Yes	I would enroll my child in a similar study or intervention program again.	6
I would recommend the treatment used to others.	Yes	I would recommend the treatment used to others.	6
I feel that the study helped improve my conversation skills.	Yes	I feel that the study helped improve my child’s conversation skills.	
I feel more confident/successful when conversing with others.	Yes	I feel that my child is more confident/successful when conversing with others.	6
I am satisfied with the outcome of the treatment for responding to nonverbal cues of boredom.	Yes	I am satisfied with the outcome of the treatment for responding to nonverbal cues of boredom	6
I am satisfied with the outcome of the treatment for asking questions.	Yes	I am satisfied with the outcome of the treatment for asking questions.	6
I am satisfied with the outcome of the treatment for repeating past conversations.	Yes	I am satisfied with the outcome of the treatment for repeating past conversations.	6
I feel that I will be able to use the skills learned in other settings and with other people.	Yes	I feel that my child will be able to use the skills learned in other settings and with other people.	6
I feel that I enjoy having conversations with others more than I did before.	Yes	I feel that my child enjoys having conversations with others more than they did before.	6
I feel that conversations are easier for me than they were before.	Yes	I feel that my child has an easier time having conversations with others than they did before.	6
<i>Pre- and Post-Treatment Videos</i>		<i>Pre- and Post-Treatment Videos</i>	
Question	Rating (“No”, “A little”, or “Yes”)	Question	Rating (scale 0-6)
There is a noticeable difference in my conversation skills in the pre- and post-treatment videos.	Yes	There is a noticeable difference in my child’s conversation skills in the pre- and post-treatment videos.	6
Seeing a video of my progress was helpful.	Yes	Seeing a video of my child’s progress was helpful.	6
I noticed improvement on how well I respond to nonverbal cues of boredom after watching the post-treatment video.	Yes	I noticed improvement on how well my child responds to nonverbal cues of boredom after watching the post-treatment video.	6
I noticed improvement on how often I asking questions after watching the post-treatment video.	Yes	I noticed improvement on the mix of statements and questions during my child’s conversations watching the post-treatment video.	6
I noticed improvement on how often I repeat past conversations after watching the post-treatment video.	Yes	I noticed improvement on how much I repeated past conversations after watching the post-treatment video.	6
My performance in the post-treatment video is similar to how I now have conversations with others outside the context of the study.	Yes	My child’s performance in the post-treatment video is comparable to how I now have conversations with others outside the context of the study.	6
<i>Comments</i>		<i>Comments</i>	
Question	Response	Question	Response
Please write anything else you want to say about things you liked and didn’t like about the study.	N/A; this section was left blank	Please write any additional comments regarding your satisfaction with the study.	N/A; this section was left blank

Table 6. Questions asked and ratings given on the social validity form issued to Participant 1 and her caregiver post-study.

Table 7 shows the questions asked and social validity ratings given in each category by Participant 2 and her caregiver. Ratings were only obtained on the pre-treatment form, as Participant 2 withdrew assent for participating in the study during the intervention phase. Participant 2 and her caregiver reported high levels of satisfaction with the goals selected as well as the intervention procedures. Participant 2's caregiver endorsed the highest rating (6) for all statements on the procedures section of the form, with the exception of a statement related to her understanding of the teaching procedures being used (given a rating of 5). The researcher met with the caregiver to review the teaching procedures again and answer any questions before proceeding to the intervention phase of the study.

<i>Participant 2 Pre-treatment Social Validity Rating</i>		<i>Participant 2 Caregiver Pre-treatment Social Validity Rating</i>	
<i>Goals</i>		<i>Goals</i>	
Question	Rating ("No", "A little", or "Yes")	Question	Rating (scale of 0-6)
The researcher listened to me about what my goals for conversation skills were.	Yes	I feel included in the selection of goals for my child.	6
I feel that overall, the goals selected to work on are important to me.	Yes	I feel that overall, the goals selected to work on with my child are important.	6
It's important to me to work on my conversation skills.	Yes	I feel that conversation skills are an important skillset for my child to work on.	6
I feel that responding to nonverbal cues of boredom is an important skill to work on.	Yes	I feel that responding to nonverbal cues of boredom is an important skill to work on.	6
I feel that staying on topic during a conversation is an important skill to work on.	Yes	I feel that using more statements during a conversation is an important skill to work on.	6
I feel that showing interest in others during a conversation is an important skill to work on.	Yes	I feel that avoiding repeating past conversations is an important skill to work on.	6
<i>Intervention</i>		<i>Intervention</i>	
Question	Rating ("No", "A little", or "Yes")	Question	Rating (scale 0-6)
I understand how I will be taught the conversation skills.	Yes	The researcher explained the treatment procedures to me in a way that I understood.	5
How I will be taught conversation skills is appropriate based on my goals.	Yes	The methods that will be used to teach my child conversation skills are appropriate based on their goals.	6
How I will be taught conversation skills is appropriate for me and my needs.	Yes	The methods that will be used to teach my child conversation skills are appropriate based on my child's needs.	6
<i>Comments</i>		<i>Comments</i>	
Question	Response	Question	Response
Please write anything else about things you liked and didn't like about the initial assessment.	N/A; this section was left blank	Please write any additional comments regarding your satisfaction with the initial assessment portion of the study.	N/A; this section was left blank

Table 7. Questions asked and ratings given on the social validity form issued to Participant 2 and her caregiver pre-treatment.

Discussion

We evaluated the extent to which BST and in-vivo training promoted the generalization of conversation skills to novel conversation partners with two adolescents with autism. First, we identified conversation skills deficits using a semi-structured direct assessment protocol based on Hood and colleagues' (2017) procedures. Then, participants experienced BST (Participants 1 and 2) and in-vivo teaching (Participant 2). The impact of the teaching procedures on the generalization of taught skills across three novel conversation partners was evaluated using a multiple probe design across behaviors. We found that BST may have been sufficient to promote generalization of conversation skills to novel conversation partners for one participant, but that BST alone was not sufficient to promote generalization for the second participant.

Participant 1

Participant 1 displayed rapid acquisition (Figure 2, 3) and generalization (Figure 1) during the study. Additionally, the target behaviors asking questions and repeating past conversation seemed to improve in the baseline, particularly after BST was introduced for responding to nonverbal cues of boredom. It seems highly likely that Participant 1 had prerequisite skills that may have contributed to the quick acquisition and generalization of targets. At the start of the study, Participant 1 was currently receiving ABA therapy with a local provider. She and her caregivers reported that social skills had previously been targeted with this provider, but that it was not a current target. Participant 1 also reported that she had *not* previously received training directly targeting the behaviors

selected for the study. However, her history of skill acquisition programming in this area may have impacted results.

Additionally, Participant 1's affect during the study was overall highly positive; she often initiated social interaction with the training confederate during break periods and consistently showed signs of happiness (e.g., smiling and laughing) during sessions. She and her caregivers frequently made statements during and after sessions regarding high social validity for the goals, procedures, and outcomes. This suggested that the high scores reported on the social validity forms issued post-assessment and at the end of the study were valid and that the participant had high levels of motivation to participate in the study. The results of the post-hoc analysis conducted on Participant 1's overall affect also suggests high levels of social validity and overall satisfaction with the study

Overall, based on the limited experimental control observed for Participant 1's data, we cannot confidently say the extent to which the treatment packages were responsible for improvements in responding observed during baseline for the second and third target behaviors. It is possible that training one skill produced improvements in the others, suggesting that the skills included as targets in this study were at least somewhat functionally interdependent. For example, perhaps Participant 1 repeating past conversations was at least in part a function of a deficit in being able to come up with or introduce new topics during a conversation. Therefore, teaching Participant 1 to change the topic during a conversation reduced the frequency in which she was observed repeating past conversation topics.

Participant 2

In comparison, Participant 2 required more trials to reach mastery criterion during BST and in-vivo training (Figure 5), and neither teaching procedure was sufficient to produce generalization for the first target behavior (responding to nonverbal cues of boredom/being annoyed; Figure 4) across conversation partners. Participant 2 appeared far less motivated than Participant 1, based on her general affect and behavior during sessions (see Figures 6 and 7). This may have contributed to the higher number of trials needed to meet mastery during training phases.

During training, Participant 2 was overall more accurate and required half as many trials to reach mastery criterion in the BST phase compared to the in-vivo teaching phase. This may have been related to poor social validity of the procedures and goals of the intervention for this participant; in general, the in-vivo teaching trials (10 minutes each) and assessment probes (20 minutes each) were longer than the amount of time it took to complete a 10-trial block during the BST phase. While all sessions were the same length (2 hours long), the participant likely experienced more break periods during the BST training phase.

However, the difference in accuracy and the rate of mastery may also suggest more structured interventions (e.g., BST) are better suited for skill acquisition for this participant. Additional components to the procedure (e.g., multiple exemplar training, contrived reinforcement contingencies, etc.) may have been necessary to transfer stimulus control of taught skills from highly contrived teaching procedures (BST) to less structured assessment and training procedures (the semi-structured direct assessment and in-vivo teaching). The purpose of introducing in-vivo teaching with the primary

confederate associated with the assessment probes was to investigate this further by training with multiple exemplars (i.e., the training confederate *and* Confederate 1). The incomplete data set obtained for Participant 2 limits the extent to which this can be evaluated.

Participant 2 did not have a history of previous training in social skills at the start of the study; however, she began attending a social skills group 3 days a week during the modeling phase of BST for the first target behavior. Participant 2 and her caregiver reported that the goals of the social skills group were different than what was being targeted in the study.

Limitations and Future Research

Undoubtedly, a limitation of the current study that is important to note is the small number of participants. In addition, the carryover observed for the second and third target behaviors for Participant 1 in baseline and the incomplete data set for Participant 2 are further limitations. The small sample size combined with the generalization of training effects across behaviors for Participant 1 and incomplete data set for Participant 2 severely limits the experimental control for this study, thus preventing us from drawing a valid inference regarding the extent to which BST and in-vivo training promoted the generalization of conversation skills with either participant. Future research should recruit a larger number of participants to account for the attrition that may occur during the study. Future researchers should also be cautious when selecting behaviors included in a multiple-probe across behaviors design, to select behaviors that are independent of each other.

Future research should also consider expanding on the assessment of social skills using procedures similar to the direct assessment procedures in this study and the Hood et al. 2017 study. The direct assessment protocol may address several limitations present in common assessment procedures currently in the literature. First, the assessment procedure is highly flexible, allowing for a wide variety of conversation skills to be assessed. Many of the skills targeted during the initial assessment in the current study were selected collaboratively with the caregiver and participant, which may have positively impacted the social validity of the goals selected.

The procedures were also more naturalistic than other direct assessments. Moving away from the highly protocolized approach that is generally taken in the previous literature may allow researchers and clinicians to get a sample more representative of the learner's repertoire. The full extent to which these assessment procedures have higher external validity needs to be further evaluated in future research.

The in-vivo teaching protocol is also more flexible and less contrived compared to other methods used in the literature. This may enhance contextual responding and discrimination of when to use certain skills, thus addressing issues related to the lack of generalization observed with social skills interventions for this population. Many other studies focused on more contrived types of training procedures, though empirically supported (e.g., video modeling, BST, etc.), continue to either report poor generalization outcomes or do not report on generalization at all. Again, the small number of participants and the results of this study limit the extent to which it can be said that in-vivo teaching is more likely to produce robust generalization outcomes.

There are further limitations of this study important to note. First, the implementation of the assessment, data collection, and treatment components may be time intensive compared to some other social skills assessments and treatments in the literature. In the case of Participant 1, the procedures, while time intensive, appeared to be effective. This may support the notion that more intensive protocols are required for repertoires such as conversation skills given the complex and nuanced nature of such a repertoire. However, the procedures were more so time intensive with Participant 2 without as robust effects. Future research could focus on ways to make this process more efficient.

Second, while the direct assessment is run loosely compared to other more structured, rigid assessments (such as ones that are scripted), the assessment still may not effectively emulate the natural environment and conversations may have felt contrived to the participants. This could have resulted in reactivity from participants during the assessment procedure.

Part of the rationale of showing the pre- and post-treatment video samples at the end of the study to the participant and caregiver as part of the social validity procedure was to attempt to evaluate the external validity of the assessment procedures. Some of the questions asked on the post-study social validity survey directly asked the participant and their caregiver to report on the extent to which the participant's performance during the assessment condition of the study is similar to how the participant has conversations in other settings. Participant 1 and her caregiver reported that the videos had high external validity. However, given the questionable validity of the social validity measures used in the study (see below), future research is needed on whether this method could be used to

evaluate the external validity of the assessment procedures. Future research could also investigate the use of more inclusive settings.

A third limitation to note is that the direct assessment procedures used in this study are not structured in a way that teases out if poor performance from a participant is due to a motivational issue, poor stimulus control, or a skills deficit. It is often assumed in the social skills literature that poor performance is the result of a skills deficit, but that may not always be the case. Differences in motivation and history of previous social skills training between participants 1 and 2 seemed to have been factors that may have impacted their performance during the study. Identifying factors such as these may be important in informing how to structure teaching procedures. Future research should focus on structuring the assessment such that all potential variables contributing to poor performance are investigated.

The data obtained on Participant 2's behavior and general affect during sessions (Figures 7 and 8) suggest that the social validity measures used in the study may not be a valid measure of social validity for this participant. While she reported high levels of satisfaction on the social validity survey (see Table 9), the frequent and consistent instances of a negative affect and indicators of boredom, and her withdrawal of assent during the in-vivo training with Confederate 1 phase indicate that the goals, procedures, and outcomes were not socially significant for her.

The lack of validity of these measures is undoubtedly a limitation to the study, and points to an important area of future research. The procedures used to assess social validity in this study were based off how our field typically measures social validity; a recent review of social validity practices found that rating scales and questionnaires were

the most common methods (Ferguson et al., 2019). The issue of the validity of current ways to assess social validity has been a discussion among behavior analysts from the start; Wolf (1978) notes that the subjective nature of this data may impact the validity of the data. However, Wolf (1978) also emphasizes the importance of collecting this data, and Baer, Wolf, and Risley (1978) further assert that consumer opinions on what is socially significant are critical to adhering to the “applied” tenant of ABA. Thus, future research should continue expanding on the use of this technology and explore other methods of assessing social validity that may be more valid and sensitive.

Given the apparent invalidity of the social validity measures used in the study, the descriptive data collected on Participants 1 and 2’s affect and behavior during session were the primary means through which decisions were made regarding whether “assent” was truly given, and if Participant 2 should continue participating in the study. There appears to be limited guidance offered in the current literature on assent procedures, despite the BACB ethics code explicitly stating that behavior analysts are responsible for obtaining assent from participants when applicable (Behavior Analyst Certification Board, 2020). Morris, Detrick, and Peterson (2021) conducted a literature review on assent practices in behavior analysis and noted that many of the articles included in the review provided little to no detail on their assent procedures.

The data obtained from Participant 2 also suggests that assent should not be assumed based on an individual’s verbal report of assent. For instance, in some circumstances there is a possibility of “coerced assent” due to complexities related to assent and guardianship (Morris et al., 2021). Additionally, given the wide range of communication skills in individuals with ASD, assent may look different for each

participant. Therefore, future research should further evaluate methods for continually assessing and reaffirming assent with a wide range of individuals with ASD. Further, researchers studying assent should consider identifying and operationally defining indicators of assent and the withdrawal of assent based on each individual participant. There are a wide range of topographies a learner may potentially engage in that indicate assent and non-assent. Measures such as caregiver interview and direct observation may help identify valid measures of assent and withdrawal of assent that are functional for each learner.

Summary

This study was an extension and replication of Hood and colleagues' (2017) method of assessing and treating social skills and aimed to evaluate the impact of in-vivo training on generalization of conversation skills to novel conversation partners. Our results suggested that BST may have been sufficient to promote generalization of conversation skills to novel conversation partners for one participant, but that BST alone was not sufficient to promote generalization for the second participant. Three major limitations (our small sample size, generalization of training effects across behaviors for Participant 1, and incomplete data set for Participant 2) prevents us from drawing a valid inference regarding the extent to which BST and in-vivo training promoted the generalization of conversation skills with either participant. However, the post-hoc analysis of each participant's affect throughout the course of the study is a novel contribution to the literature and may be a direction for future research.

References

- Baer, Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *20*(4), 313–327.
<https://doi.org/10.1901/jaba.1987.20-313>
- Beaulieu, L., Hanley, G. P., & Santiago, J. L. (2014). Improving the conversational skills of a college student with peer-mediated behavioral skills training. *The Analysis of verbal behavior*, *30*(1), 48-53.
- Behavior Analyst Certification Board. (2020). *Ethics code for behavior analysts*.
<https://bacb.com/wp-content/ethics-code-for-behavior-analysts/>
- Chezan, L. C., Dragow, E., & Grybos, E. M. (2020). Conversation skills and self-initiated interactions in young adults with autism and intellectual disability. *Research in Autism Spectrum Disorders*, *75*, 101554.
- Custer, T. N., Stiehl, C. M., & Lerman, D. C. (2020). Outcomes of a practical approach for improving conversation skills in adults with autism. *Journal of Applied Behavior Analysis*.
- Davis, K. M., Boon, R. T., Cihak, O. E., & Fore, III (2010). Power cards to improve conversational skills in adolescents with Asperger syndrome. *Focus on Autism and Other Developmental Disabilities*, *25*, 12-22 doi:10.1177/10883S7609354299
- Dotson, W. H., Leaf, J. B., Sheldon, J. B., & Sherman, J. A. (2010). Group teaching of conversational skills to adolescents on the autism spectrum. *Research in Autism Spectrum Disorders*, *4*(2), 199-209.
- Ferguson, Cihon, J. H., Leaf, J. B., Van Meter, S. M., McEachin, J., & Leaf, R. (2019). Assessment of social validity trends in the journal of applied behavior analysis. *European*

Journal of Behavior Analysis, 20(1), 146–157.

<https://doi.org/10.1080/15021149.2018.1534771>

Freeman, J.B., & Cronin, P. (2017). Standardized assessment of social skills in autism spectrum disorder. *Handbook of social skills and autism spectrum disorder*. (83-96).

Gillis, J. M., Callahan, E. H., & Romanczyk, R. G. (2011). Assessment of social behavior in children with autism: The development of the behavioral assessment of social interactions in young children. *Research in Autism Spectrum Disorders*, 5(1), 351-360.

Doi:10.1016/j.rasd.2010.04.019

Hood, S.A., Beauchesne, B.M., Fahmie, T.A., & Go, A. (2021). Descriptive assessment of conversational skills: towards benchmarks for young adults with social deficits. *Journal of Applied Behavior Analysis*, 1-20. Doi:10.1002/jaba.831

Hood, S. A., Luczynski, K. C., & Mitteer, D. R. (2017). Toward meaningful outcomes in teaching conversation and greeting skills with individuals with autism spectrum disorder. *Journal of Applied Behavior Analysis*, 50(3), 459-486. Doi:10.1002/jaba.388

Horner, D.R., & Baer, D.M. (1978). Multiple-probe technique: a variation of the multiple baseline. *Journal of Applied Behavior Analysis*, 11(1), 189-196.

Lerman, D. C., White, B., Grob, C., & Laudont, C. (2017). A clinic-based assessment for evaluating job-related social skills in adolescents and adults with autism. *Behavior Analysis in Practice*, 10(4), 323-336. Doi:10.1007/s40617-017-0177-9

Matson, J. L., & Wilkins, J. (2007). A critical review of assessment targets and methods for social skills excesses and deficits for children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 1(1), 28-37.

- Morris, C., Detrick, J. J., & Peterson, S. M. (2021). Participant assent in behavior analytic research: Considerations for participants with autism and developmental disabilities. *Journal of Applied Behavior Analysis, 54*(4), 1300–1316. <https://doi.org/10.1002/jaba.859>
- Nuernberger, J. E., Ringdahl, J. E., Vargo, K. K., Crumpecker, A. C., & Gunnarsson, K. F. (2013). Using a behavioral skills training package to teach conversation skills to young adults with autism spectrum disorders. *Research in Autism Spectrum Disorders, 7*, 441–447. <https://doi.org/10.1016/j.rasd.2012.09.004>
- Parker, A. N., Sidener, T.M., DeBar, R. M., Reeve, S. A., & Reeve, K. F. (2022). Identifying and training skill acquisition protocol development: A preliminary investigation. *Behavior Analysis in Practice*
- Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2013). Teaching practitioners to conduct behavioral skills training: A Pyramidal Approach for Training Multiple Human Service Staff. *Behavior analysis in practice, 6*(2), 4–16. doi:10.1007/BF03391798
- Ratto, A. B., Turner-Brown, L., Rupp, B. M., Mesibov, G. B., & Penn, D. L. (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.
- 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*(2), 149-166. doi:10.1007/s10803-009-0842-0

- Sansosti, F.J., & Powell-Smith, K.A. (2008). Using computer-presented social stories and video models to increase the social communication skills of children with high-functioning autism spectrum disorders. *Journal of Positive Behavior Interventions*, *10*(3), 162-178. doi:10.1177/1098300708316259
- Tse, J., Strulovitch, J., Tagalakis, V., Meng, L., & Fombonne, E. (2007). Social skills training for adolescents with asperger syndrome and high-functioning autism. *Journal of Autism and Developmental Disorders*, *37*(10), 1960-1968. Doi:10.1007/s10803-006-0343-3
- Verhoeven, E. W. M., Smeekens, I., & Didden, R. (2013). Brief report: Suitability of the social skills performance assessment (SSPA) for the assessment of social skills in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(12), 2990-2996. doi:10.1007/s10803-013-1843-6
- Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*, 203–214. <http://dx.doi.org/10.1901/jaba.1978.11-203>.

Appendix A

Sample Confederate Reference Sheet: Direct Assessment

Date: _____

Participant ID#: _____

Confederate: _____

Train or probe? _____

Greeting: _____

Give a compliment: _____

Change conversation topic: _____

Ask a question: _____

Issue Non-verbal social cue: _____

Issue Verbal cue: _____

Interruption of client verbal behavior: _____

Introduce a complicated topic: _____

Introduce a taboo or inappropriate topic: _____

Goodbye/closing: _____

Appendix B

Caregiver Initial Interview

Date: _____

Participant ID#: _____

Interviewer: _____

Start with questions 1-4. Ask follow up questions for more detail if needed.

1. Tell me about what your child struggles with during conversations with others.

2. Tell me about things you feel your child should do during conversations with others that they do not do.

3. Tell me about things you feel your child does do during conversations that they shouldn't do or do less of.

4. Tell me about things your child does well during conversations with others.

5. Tell me what your child enjoys talking about

6. Tell me what your child does not like to talk about.

On a scale of 1-7, 1 being does not do or needs a lot of help with to 7 being does well in a conversation, rate your child's performance in the following skill areas:

7. Ask and answer questions
8. Initiate conversation
9. Interrupting (both interrupting others and responding to interruptions?)
10. Give or accept compliments
11. Stay on topic
12. Pick up on verbal social cues
13. Pick up on nonverbal social cues

Please list the top five goals you feel are most important for your child's conversation skills.

- 1.
- 2.
- 3.
- 4.
- 5.

Eligibility to Participate in Study

Ask the caregiver each of the questions below to determine if the child is eligible to participate in the study.

1. Is the child able to engage in back-and-forth conversation at least for a brief period of time? Y / N
2. Does the child engage in challenging behavior that would inhibit skill acquisition (e.g., aggression, self-injurious behavior, property destruction, elopement, flopping, etc.) Y / N
3. Is the child able to engage in a work task for up to 20 minutes without taking a break? Y / N

Appendix C

Participant Initial Interview

Date: _____

Participant ID#: _____

Interviewer: _____

Start with questions 1-4. Ask follow up questions for more detail if needed.

1. Tell me about what you feel you struggle with during conversations with others.
2. Tell me about things you think you should do during conversations with others that you do not do.
3. Tell me about things you feel you do during conversations that you shouldn't do or do less of.
4. Tell me about things you do well during conversations with others.
5. Tell me what you like to talk about.
6. Tell me what you don't like to talk about.

On a scale of 1-7, 1 being needs help with to 7 being does well in a conversation, rate your performance in the following skill areas:

7. Ask and answer questions
8. Initiate conversation
9. Interrupting (both interrupting others and responding to interruptions?)
10. Give or accept compliments
11. Stay on topic
12. Pick up on verbal social cues
13. Pick up on nonverbal social cues

Please list the top five things you would like to be better at during conversations.

- 1.
- 2.
- 3.
- 4.
- 5.

Appendix D

Participant Initial Observation Checklist

Date: _____

Participant ID#: _____

Interviewer: _____

Eligibility to Participate in Study

During interactions and observation of child, answer questions below to determine if the child is eligible to participate in the study.

1. Is the child able to speak in full sentences? (note they must demonstrate this at least once)
Y / N
2. Is the child able to engage in back-and-forth conversation at least for a brief period of time? Y / N
3. Does the child engage in challenging behavior that would inhibit skill acquisition (e.g., aggression, self-injurious behavior, property destruction, elopement, flopping, etc.) Y / N
4. Is the child able to engage in a work task for up to 20 minutes without taking a break? Y / N

Semi-structured observation

Note the participant's behavior in the following domains during each prompt below and note any other potential target behaviors to assess:

Nonverbal behaviors:

- Eye contact
- Distance from speaker
- Presence of any distracting nonvocal behavior (e.g. use of cell phone)
- Listener behavior (e.g., nodding head, "mhmm", etc.)
- Others:

Verbal behaviors:

- Ask and answer questions
- Initiate conversation
- Interrupting (both interrupting others and responding to interruptions?)
- Give or accept compliments
- Stay on topic
- Pick up on verbal social cues
- Pick up on nonverbal social cues
- Others:

Conversation prompts

Instruct participant and parent to talk about the following topics, each for five minutes:

- Instruct participant and parent to talk about a topic preferred by participant.
- Instruct participant and parent to talk about a topic preferred by the parent

Appendix E

Baseline/Probe Condition Procedural Integrity Data

Date: _____

Participant ID#: _____

Confederate: _____

Observer: _____

Checklist

Endorse Y if the confederate implemented each step below, endorse N if the confederate did not implement a step correctly at least once during the session.

- Did the confederate start the condition by entering the room with the participant already seated and allow 10 seconds for the participant to initiate a greeting prior to starting the conversation? Y / N
- Did the confederate end the conversation after 20 minutes or after the participant says goodbye? Y / N
- Did the confederate avoid providing any feedback to the participant during the conversation about whether or not their responses were correct? Y / N
- Did the confederate keep all statements simple and relatively uncomplicated (with the exception of any social presses that require otherwise)? Y / N

Score (number of Ys divided by 5 x 100): ____%

Social Presses

For each social press, complete the table below.

Press	Delivered press correctly (per definition of social press)?	Allow 10 seconds to respond?	Notes
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
	Y / N	Y / N	
Score (number of Ys divided by total number of social presses x 100)	___/___ (___%)		

Appendix F

BST Treatment Integrity Data

Date: _____

Participant ID#: _____

Confederate: _____

Observer: _____

Treatment phase (circle): Discrimination training / Role play

Checklist

Endorse Y if the confederate implemented each step below, endorse N if the confederate did not implement a step correctly at least once during the session.

- Did confederate give summary of rationale, explanation of why to engage in skill and context to engage in skill at beginning of session (full explanation if 1st session, brief summary for subsequent sessions)? Y / N
- Did confederate run sessions in ten trial blocks with brief breaks in between trial blocks? Y / N

For modeling sessions ONLY (leave blank if role-play session):

Trial (COR or INC model)	Delivered S ^D Correctly?	Allow 5 seconds to respond?	Deliver appropriate consequence?	Notes
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
Score (number of Ys divided by total number of social presses x 100)	$\frac{\quad}{\quad}$ ($\frac{\quad}{\quad}$ %)			

For role-play sessions ONLY (leave blank if modeling session)

Trial (COR or INC model)	Delivered S ^D Correctly?	Allow 5 seconds to respond?	Deliver appropriate consequence?	Notes
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
	Y / N	Y / N	Y / N	
Score (number of Ys divided by total number of social presses x 100)	$\frac{\quad}{\quad}$ ($\quad\%$)			

Appendix G

In-vivo Training Condition Procedural Integrity Data

Date: _____

Participant ID#: _____

Confederate: _____

Observer: _____

Checklist

Endorse Y if the confederate implemented each step below, endorse N if the confederate did not implement a step correctly at least once during the session.

- Did the confederate start the condition by entering the room with the participant already seated and allow 10 seconds for the participant to initiate a greeting prior to starting the conversation? Y / N
- Did the confederate end the conversation after 10 minutes? Y / N
- Did the confederate keep all statements simple and relatively uncomplicated (with the exception of any social presses that require otherwise)? Y / N

Score (number of Ys divided by 5 x 100): ____%

Social Presses

For each social press, complete the table below.

Press	Delivered press correctly (per definition of social press)?	Allowed 10 seconds to respond?	Delivered appropriate prompt card within five seconds of error?	Notes
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
	Y / N	Y / N	Y / N / N/A	
Score (number of Ys divided by total number of social presses x 100)	___/___ (___%)			

Appendix H

SAMPLE Pre-treatment Social Validity Form (Caregivers)

Date: _____

Participant ID#: _____

Please give a rating for each statement below by circling one number between 0-6 based on how much you agree with the statement.

0- Strongly disagree

1- Disagree

2- Somewhat disagree

3- Neither agree nor disagree

4- Somewhat agree

5- Agree

6- Strongly agree

Goals

I feel included in the selection of goals for my child.

0 1 2 3 4 5 6

I feel that overall, the goals selected to work on with my child are important.

0 1 2 3 4 5 6

I feel that conversation skills are an important skill set for my child to work on.

0 1 2 3 4 5 6

I feel that (target #1) is an important conversation skill to work on.

0 1 2 3 4 5 6

I feel that (target #2) is an important conversation skill to work on.

0 1 2 3 4 5 6

I feel that (target #3) is an important conversation skill to work on.

0 1 2 3 4 5 6

Intervention

The researcher explained the treatment procedures to me in a way that I understood.

0 1 2 3 4 5 6

The methods that will be used to teach my child conversation skills are appropriate based on their goals.

0 1 2 3 4 5 6

The methods that will be used to teach my child conversation skills are appropriate based on my child's needs.

0 1 2 3 4 5 6

Comments

Please write any additional comments regarding your satisfaction with the initial assessment portion of the study below.

Appendix I

SAMPLE Pre-treatment Social Validity Form (Participants)

Date: _____

Participant ID#: _____

Below each statement, circle the face that matches how much you agree with the statement.

Goals

The researcher listened to me about what my goals for conversation skills were.

No

A little

Yes



I feel that overall, the goals selected to work on are important to me.

No

A little

Yes



It's important to me to work on my conversation skills.

No

A little

Yes



I feel that (target #1) is an important conversation skill to work on.

No

A little

Yes



I feel that (target #2) is an important conversation skill to work on.

No *A little* *Yes*



I feel that (target #3) is an important conversation skill to work on.

No *A little* *Yes*



Intervention

I understand how I will be taught the conversation skills.

No *A little* *Yes*



How I will be taught conversation skills is appropriate based on my goals.

No *A little* *Yes*



How I will be taught conversation skills is appropriate for me and my needs.

No *A little* *Yes*



Comments

Please write anything else about things you liked and didn't like about the initial assessment.

Appendix J

SAMPLE Post-treatment Social Validity Form (Caregivers)

Date: _____

Participant ID#: _____

Please give a rating for each statement below by circling one number between 0-6 based on how much you agree with the statement.

0- Strongly disagree

1- Disagree

2- Somewhat disagree

3- Neither agree nor disagree

4- Somewhat agree

5- Agree

6- Strongly agree

Effects of Intervention

My child enjoyed meeting with the researchers during the study.

0 1 2 3 4 5 6

I would enroll my child in a similar study or intervention program again.

0 1 2 3 4 5 6

I would recommend the treatment used to others.

0 1 2 3 4 5 6

I feel that the study helped improve my child's conversation skills.

0 1 2 3 4 5 6

I feel that my child is more confident/successful when conversing with others.

0 1 2 3 4 5 6

I am satisfied with the outcome of the treatment for (target #1).

0 1 2 3 4 5 6

I am satisfied with the outcome of the treatment for (target #2).

0 1 2 3 4 5 6

I am satisfied with the outcome of the treatment for (target #3).

0 1 2 3 4 5 6

I feel that my child will be able to use the skills learned in other settings and with other people.

0 1 2 3 4 5 6

I feel that my child enjoys having conversations with others more than they did before.

0 1 2 3 4 5 6

I feel that my child has an easier time having conversations with others than they did before.

0 1 2 3 4 5 6

Pre and Post Treatment Videos

There is a noticeable difference in my child's conversation skills in the pre- and post-treatment videos.

0 1 2 3 4 5 6

Seeing a video of my child's progress was helpful.

0 1 2 3 4 5 6

I noticed improvement on (target #1) after watching the post-treatment video.

0 1 2 3 4 5 6

I noticed improvement on (target #2) after watching the post-treatment video.

0 1 2 3 4 5 6

I noticed improvement on (target #3) after watching the post-treatment video.

0 1 2 3 4 5 6

My child's performance in the post-treatment video is comparable to how my child now has conversations with others outside of the context of the study.

0 1 2 3 4 5 6

Comments

Please write any additional comments regarding your satisfaction with the study below.

Appendix K

SAMPLE Post-treatment Social Validity Form (Participants)

Date: _____

Participant ID#: _____

Below each statement, circle the face that matches how much you agree with the statement.

Effects of Intervention

I enjoyed meeting with the researchers during the study.

*No**A little**Yes*

I would participate in a similar study or intervention program again.

*No**A little**Yes*

I would recommend the treatment used to others.

*No**A little**Yes*

I feel that the study helped improve my conversation skills.

*No**A little**Yes*

I feel more confident/successful when conversing with others.

No

A little

Yes



I am satisfied with the outcome of the treatment for (target #1).

No

A little

Yes



I am satisfied with the outcome of the treatment for (target #2).

No

A little

Yes



I am satisfied with the outcome of the treatment for (target #3).

No

A little

Yes



I feel that I will be able to use the skills learned in other settings and with other people.

No

A little

Yes



—

I feel that I enjoy having conversations with others more than I did before.

No

A little

Yes



I feel that conversations are easier for me than they were before.

No

A little

Yes



—

Pre and Post Treatment Videos

There is a noticeable difference in my conversation skills in the pre- and post-treatment videos.

No

A little

Yes



Seeing a video of my progress was helpful.

No

A little

Yes



I noticed improvement on (target #1) after watching the post-treatment video.

No

A little

Yes



I noticed improvement on (target #2) after watching the post-treatment video.

No

A little

Yes



I noticed improvement on (target #3) after watching the post-treatment video.

No

A little

Yes



My performance in the post-treatment video is similar to how I now have conversations with others outside of the context of the study.

No

A little

Yes



Comments

Please write anything else you want to say about things you liked and didn't like about the study.