Evaluating Change in Social Skills in High-Functioning Adults With Autism Spectrum Disorder Using a Laboratory-Based Observational Measure

Susan W. White, PhD[1], Angela Scarpa, PhD[1], Caitlin M. Conner, MS[1], Brenna B. Maddox, MS[1], and Saray Bonete, PhD[2]

Abstract

Despite the impact of social disability on the lives of people with Autism Spectrum Disorders (ASDs), we know little about how to assess clinical improvement in this domain. This is a preliminary study of the potential utility and sensitivity of a novel observational rating system, the Contextual Assessment of Social Skills (CASS), as a treatment outcome measure with cognitively unimpaired adults with ASD. Five cognitively unimpaired adults with ASD completed the CASS, before and after a group social skills intervention. Based on ratings made by evaluators masked to assessment (pre-treatment or post-treatment), reliable change indices were computed to assess improvement. Four participants demonstrated significant improvement in conversational involvement, two initiated significantly more topic changes, and one asked more questions. Laboratory-based observational measures, such as the CASS, may be useful in clinical trials for adults with ASD, though further evaluation with larger samples is required.

The fundamental area of impairment characteristic of people with Autism Spectrum Disorder (ASD), including those who are higher functioning cognitively (i.e., termed "HFASD"), is severe and chronic social disability. Severity of social deficits in people with ASD is associated with greater functional impairment and poorer quality of life (Segrin & Givertz, 2003), underscoring the importance of identifying effective treatment approaches. Although far from unequivocal, the extant research on interventions for social deficits in people with ASD is promising (White & Maddox, 2013). Treatment gains, however, are typically modest and often not sustained long-term beyond treatment (Bellini, Peters, Benner, & Hopf, 2007). Progress in developing and disseminating effective treatment approaches, including our ability to make valid comparisons among approaches and programs, has been hindered by a lack of sensitive and valid measures with which to assess small yet meaningful improvements in social functioning (Rao, Beidel, & Murray, 2008; White, Koenig, & Scahill, 2007).

There exists no "gold standard" measure for assessing outcomes with social skills interventions for clients with ASD (Cunningham, 2011; Lord et al., 2005; Wolery & Garfinkle, 2002), although many different tools are available (Merrell, 2001). Based on a review of the extant

literature on evaluation of social functioning in people with ASD following treatment (see Table 1), it can be seen that the most popular tools are rating scales, followed by behavioral observations, which can be either laboratory-based or naturalistic. Rating scales, however, are not without limitations. They are sensitive to respondent bias, difficult to use in a masked fashion (in which the rater is unaware of treatment or control condition) because the respondent is typically the parent or participant, and can be very context dependent, such that ratings vary widely for the same individual (e.g., at home and at school; Merrell, 2001). In addition, with the exception of the second edition of the *Social Responsiveness Scale* (SRS-2; Constantino & Gruber, 2012), recently expanded to be used with adults, none of the rating scales described in Table 1 were designed for use with adults. Most available rating scales are only

---

[1]Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, USA
[2]Department of Psychology, Francisco Vitoria University, Madrid, Spain

Corresponding Author:
Susan W. White, Child Study Center, Virginia Polytechnic Institute and State University, 460 Turner St., Blacksburg, VA 24060, USA.
Email: sww@vt.edu

Table 1. Measures of Social Functioning Sensitive to Change.

Measure	Author(s)	Age range	Description
Rating scales			
SSRS	Gresham and Elliott (1990)	3 forms: Pre-school (age 3–5); Elementary (Grades K–6); Secondary (Grades 7–12)	40–57 items for adult raters and 34–39 item Likert-type scale for students. Comprehensive multi-rater (parent, teacher, self) assessment. Measures teacher–student relations, peer interactions, and academic performance
SSIS-RS	Gresham and Elliott (2008)	Age 3–Grade 12	Approximately 140 items per form (depends on form and age of child). Updated version of the SSRS with additional subscales and an Intervention Guide to provide appropriate social interventions after a Performance Screening and measure progress
SRS-2	Constantino and Gruber (2005, 2012)	Ages 4 through adult	65-item Likert-type scale filled out by parent/caregiver, teacher, or self. Assesses characteristics of ASD across five domains
Children's Social Behavior Questionnaire	Luteijn, Jackson, Volkmar, & Minderaa (1998)	Ages 4–18	96-item Likert-type scale to be completed by parent/caregiver. Includes measures of communication, social interaction, stereotypy, motor behaviors, affect, attention, sensory abnormalities, and social cues
MESSY	Matson, Rotatori, and Helsel (1983)	Ages 4–18	64-item Likert-type rating scale with teacher and self-report forms. Items consist of discrete observable behaviors rather than traits
Behavioral observation			
Lab-based behavioral observation			
ESCS	Mundy et al. (2003)	Ages 8–30 months	Videotaped observation focused on non-verbal communication skills in response to toys and the examiner. Extensive manual available. Measures of Joint Attention Behaviors, Behavioral Requests, and Social interaction behaviors
ADOS-2	Lord, Rutter, DiLavore, and Risi (2002); Lord et al. (2012)	Ages 12 months through adult	A 30–60 min semi-structured behavioral observation conducted by a trained reliable professional. Foremost used to diagnose ASD but has been found to be sensitive to change when used as an outcome measure
Naturalistic behavioral observation			
Behavior Coding Scheme	Hauck, Fein, Waterhouse, and Feinstein (1995)	School-aged children	Trained observer in naturalistic setting rates on positive, negative, low level, attention seeking initiation, and avoidance during 15-s intervals of play over a total observation time of 15 min.
POPE	Kasari, Rotheram-Fuller, and Locke (2005)	School-aged children	Time-interval behavior coding system. After 40 s watched, two blinded observers rate children's engagement with peers on the playground, frequency of initiations and responses within 20 s intervals for at least 10 min during recess or lunch play on two separate occasions

Note. SSRS = Social Skills Rating System; SSIS-RS = *Social Skills Improvement System–Rating Scale*; SRS-2 = *Social Responsiveness Scale–Second Edition;* ASD = Autism Spectrum Disorder; MESSY = Matson Evaluation of Social Skills in Youngsters; ESCS = *Early Social Communication Scale*; ADOS-2 = *Autism Diagnostic Observation Schedule–Second Edition*; POPE = Playground Observation of Peer Engagement.

appropriate for use with school-aged children, whereas impairments in social skills are usually lifelong in people with ASD (e.g., Billstedt, Gillberg, & Gillberg, 2005; Smith, Maenner, & Seltzer, 2012).

Observational assessments may arguably be more sensitive to change and be psychometrically more reliable than questionnaires. Behavioral observations are typically administered in a standardized way using codified procedures and ratings made by trained assessors. Moreover, observational assessments can be individualized to examine specific behaviors or skills that are targeted in treatment (Cunningham, 2011). Natural setting observations, in which behaviors are observed and recorded in the person's normal daily life (e.g., at school), are highly ecologically valid. The benefit of assessment in a natural setting is sometimes offset, however, by variability across observation sessions and raters, time/scheduling burden, and risk of inaccurate data due to observer reactivity and lack of social comparison data, or a normative comparison of what the observed behavior might look like (Merrell, 2001). Laboratory-based tasks overcome some of these difficulties, owing to standardized administration and the ability to exert more control than can be afforded with naturalistic observations. For example, the examiner can eliminate unexpected intrusions during the observation period and use specific probes meant to elicit certain kinds of normative or non-normative social behaviors. Laboratory-based observational assessments may provide a useful tool to determine whether social skills targeted in treatment generalize beyond the treatment context. Given the chronic nature of social impairments in ASD (Eaves & Ho, 2008; Losh et al., 2009) and the dearth of research on effective interventions for social disability in adults with ASD (Mazefsky & White, 2013; Reichow & Volkmar, 2010; Wang & Spillane, 2009), it is critical that we identify measures of social competence for adults with ASD that are reliable and valid as well as sensitive to change. Laboratory-based assessments may be especially well-suited for use with adult clients, given the previously described strengths of this approach, as well as difficulty conducting naturalistic observations with adults and the need for blinded assessment in clinical trials.

The Contextual Assessment of Social Skills (CASS; Ratto, Turner-Brown, Rupp, Mesibov, & Penn, 2011) is a recently developed laboratory-based observational measure of social functioning and conversational skills developed for adolescents and young adults with HFASD. There are several reasons why the CASS is a promising gauge for treatment response, at least with respect to specific social behaviors. Because it was developed for adults with HFASD, it pulls for behaviors often deficient, or sometimes present in excess, in this population (e.g., frequency of topic changes initiated or kinetic arousal, respectively), and it has been demonstrated that naïve raters can be trained to reliably code the CASS from videotape (Ratto et al., 2011). The

fairly structured administration (e.g., timed pauses before the confederate should initiate speech) also ensures that participants have comparable opportunities to demonstrate conversational strengths and weaknesses. Perhaps most importantly, the CASS is a conversational task in which the assessed individual (the person with ASD) interacts with two confederates (i.e., unfamiliar, opposite sex, similar-age adults). Given that social interactions with peers typically constitute the most challenging interpersonal situations for people with ASD (Paul, Orlovski, Marcinko, & Volkmar, 2009), this structure is especially relevant for evaluation of change in social dysfunction in this population.

The purpose of the present study was to examine the potential utility of the CASS as an outcome measure with adults with HFASD in the context of a small, open clinical trial using an evidence-based intervention for young adults with ASD, the Program for the Education and Enrichment of Relationship Skills for Young Adults (PEERS-YA; Laugeson & Frankel, 2014). Social competence, as indexed by the CASS, involves initiating an interaction with an unfamiliar peer, managing verbal and non-verbal conversational skills, and assessing a conversational partner's receptiveness and what to do when he or she is uninterested (i.e., not responsive). Many of these social behaviors are targeted in PEERS-YA. In addition to exploring trainability of raters and interrater reliability, we hypothesized that improvement in specific, measured social behaviors as observed in the context of the CASS would be observed post-treatment, compared with pre-treatment. We secondarily hypothesized that social adaptation, or the person's ability to recognize conversational boredom in another adult and modulate one's behavior accordingly, would improve following treatment. Given the high prevalence of problems with social anxiety in this population (Kreiser & White, 2014) and the bi-directional relationship between social anxiety and social deficits (White et al., 2013), we also sought to explore whether change in social anxiety, as measured observationally with the CASS, would correlate with change in self-reported social anxiety. Finally, we evaluated the degree to which improved CASS scores correlated with improvement on rating scales of social competence.

Method

Participants

To be eligible for the study, participants had to be between the ages of 18 and 28, have an ASD diagnosis, and have an adult in their life who could serve as their assistant in the treatment program (e.g., a parent or roommate). Diagnoses of ASD from the community were accepted, on the condition that participants provided documentation (psychological reports) supporting the diagnosis. Participants had to be free of intellectual disability, as measured by the *Kaufman*

Table 2. Demographic Information.

Participant	Age (years)	Level of schooling	Diagnosis	Total IQ	VIQ	NVIQ	SRS self-report *T*-score	SRS caregiver-report *T*-score
James	19.42	College student	PDD-NOS	105	101	107	64	78
Mark	22.00	College student	Asperger Disorder	58	75	52	64	55
Drew	25.00	Completed high school	Asperger Disorder	113	102	120	81	73
Anthony	27.33	Completed high school	Autistic Disorder	78	62	100	77	73
Arnold	26.67	Completed some college	Asperger Disorder	107	98	115	64	60

Note. VIQ = Verbal IQ; NVIQ = non-verbal IQ estimated by K-BIT; SRS = *Social Responsiveness Scale*; K-BIT = *Kaufman Brief Intelligence Test*; PDD-NOS = Pervasive Developmental Disorder–Not Otherwise Specified.

Brief Intelligence Test–Second Edition (K-BIT-2; Kaufman & Kaufman, 2004), and be free of psychopathology that would warrant more immediate or intensive clinical treatment (e.g., psychosis). All participants fell between the mild and severe range of ASD severity on the SRS-2 self-report (Constantino & Gruber, 2012). Table 2 presents demographics for each participant.

Of the five participants, all were male, self-identified as Caucasian, and were unemployed and resided with their parents. All of the adults comprising the assistant group, which runs concurrently with the groups for the adults with ASD, were the participants' biological parent(s). In the present report, pseudonyms are used to protect confidentiality. James, a 20-year-old man diagnosed with Pervasive Developmental Disorder–Not Otherwise Specified (PDD-NOS)—all diagnoses were made using *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM-IV-TR*; American Psychiatric Association [APA], 2000) criteria—was enrolled at a community college. He displayed anxiety about speaking and participating in the group, especially at the outset of the group, but appeared to become more comfortable speaking during group sessions as the group progressed. Mark, a 22-year-old man diagnosed with Asperger Disorder, also attended classes at a community college. Although he denied needing assistance with social skills, he expressed wanting to feel less anxious in social situations. Mark was engaged in, and appropriate during, the group meetings and polite with other group members and the therapists. Drew, a 25-year-old man with Asperger Disorder, reported wanting to make more friends. He readily spoke with other group members before and after group sessions. Anthony, a 26-year-old man with Autistic Disorder, struggled with social interaction and emotion regulation, specifically controlling his anger when things did not go as he anticipated. Although his verbal IQ was low (62), he was included in the group because he was motivated to participate and seemed to understand the instructions during the intake assessment. During the intervention, Anthony participated fully, but he often struggled to understand the presented material. The group leaders emailed his mother an overview of the content prior to

every session to facilitate his understanding. Arnold was a 26-year-old man with Asperger Disorder whose self-reported goal for being in the program was to improve his social skills and to "talk to others better." He was highly engaged with the other group members, initiating conversation both before and after group meetings.

Intervention: PEERS-YA

The original PEERS Intervention is a 14-week, parent-assisted social skills group intervention for adolescents with HFASD (Laugeson & Frankel, 2010). PEERS has been adapted for high-functioning young adults (PEERS-YA; Laugeson & Frankel, 2014), and two recent studies have supported the efficacy of PEERS through randomized controlled trials with adolescents (Laugeson, Frankel, Gantman, Dillon, & Mogil, 2012) and young adults (Gantman, Kapp, Orenski, & Laugeson, 2012). The PEERS-YA sessions are delivered in a small-group format, with concurrent groups for the adults with ASD and for their caregivers (usually the parents). Each session focuses on a particular lesson, or skill, related to building interpersonal relationships (e.g., organizing get-togethers, dating etiquette). PEERS-YA targets social independence in the young adult with ASD through directive teaching, rehearsal, role-play, feedback, and weekly homework. Laugeson et al. (2012) observed improvements in parent-reported ASD symptoms, social skills knowledge, and amount of time spent with others that were maintained at the 14-week follow-up assessment. Similarly, Gantman and colleagues (2012) found improved communication and decreased ASD mannerisms post-intervention in adults following completion of the PEERS-YA program. Although very promising, these studies quantify improvement solely through rating scales and questionnaires. In the present study, the intervention was implemented by two graduate student clinicians, one of whom led the young adult group, whereas the second led the assistant group. Therapists were supervised by two licensed clinical psychologists and were trained by the PEERS developers.

Measures

The following measures were administered immediately prior to starting the PEERS intervention and again following the final treatment session (approximately 14 weeks later).

CASS. An observational measure of social competence developed for adolescents and young adults with ASD, the CASS is comprised of two, 3-min role-play conversations with two similar-age, opposite-gender confederates (both of whom are unfamiliar to the participant; Ratto et al., 2011). Opposite-gender confederates are stipulated by the CASS protocol, as ability to communicate with the opposite gender is one of the primary tasks during adolescence and young adulthood. The CASS measures conversational skills across relevant dimensions and specifically focuses on the individual's ability to adapt one's behavior in response to change in the social context (i.e., when the conversational partner is no longer interested, but rather disinterested). Adults with ASD are expected to display greater stability in behavior (i.e., inflexibility in social response to the partner's change in affect and behavior) across social context. Consistent with this hypothesis, Ratto et al. (2011) found that social adaptation as measured by the CASS predicted ASD status.

After being instructed by the examiner to "act as if you have recently joined a new club or social group" (see Ratto et al., 2011, for detailed administration instructions), the participant and confederate are left alone in the room to talk to each other. In the first conversation, the confederate displays a socially interested, engaged demeanor verbally and non-verbally (i.e., Interested condition: I-CASS). In the second conversation, a different confederate displays a bored/disinterested demeanor, with minimal verbal responses and eye contact (i.e., Bored condition: B-CASS). The CASS is videotaped for later coding, and codes are assigned for each condition based on the participant's verbal and non-verbal behaviors, across four core domains (Asking Questions, Topic Changes, Overall Involvement, and Overall Quality of Rapport) and five exploratory domains (Social Anxiety, Kinesic Arousal, Vocal Expressiveness, Gestures, and Positive Affect). Ratto et al. (2011) reported that the CASS effectively discriminated between adults with and without ASD on seven of the nine domains (Gestures and Kinesic Arousal did not differentiate). Group differences were also apparent in the *Social Adaptation Index* (SAI), or the person's ability to adjust his or her own behavior when the conversational partner is disinterested.

Prior to the present study's launch, the developer of the CASS (A. Ratto) conducted on-site training to explain the measure's development, administration, and scoring. Following this training, six raters were trained to reliability

(i.e., >80% agreement with gold standard ratings) on scoring using the set of six training vignettes from the original study (Ratto et al., 2011). Two additional raters were later trained to the reliability threshold using the six original training videos and two additional videos (with adults with ASD). Five undergraduate students and four clinical psychology graduate students served as CASS confederates (all females, as the PEERS-YA participants were all males). The confederate training lasted for approximately 1 hr, followed by dyad practice among the confederates. They were trained according to each context (i.e., interested or bored), to use appropriately timed pauses and to manipulate eye contact, facial affect, posture, and gestures, as well as to minimize social initiation and to make brief statements in the bored context as described by Ratto et al. (2011) and in the training protocol. All confederates practiced both contexts with the lead author (omitted for masked review) before completing a CASS session.

To examine change in social behavior measured by the CASS, we first examined change (pre- to post-treatment) in the social behaviors assessed only in the Interested condition, for each of the nine separate domains coded, as we were most interested in examining change after treatment across the specific social behaviors that were addressed and practiced during the group intervention. In addition, the confederate's behavior during the Interested condition of the CASS is a better representation of typical social discourse (Le Poire & Yoshimura, 1999). The I-CASS Total Score was also calculated to examine construct validity via comparison with the rating scales implemented. The Total Score is comprised of the sum of the four core domains (Asking Questions, Topic Changes, Overall Involvement, and Overall Quality of Rapport) after item score conversion to standard scores. The ability to modulate one's own behavior on the basis of a conversational partner's behavior is arguably a higher level social skill, and therefore, it should be considered an independent score provided by the CASS. The SAI was calculated using the approach of the CASS authors (Ratto et al., 2011) such that rapport was reverse-scored (because a decline in rapport is normative in the Bored compared with the Interested context), then all four core items were converted to standard scores and summed independently for each condition, Interested and Bored. The SAI is the difference between the sum of these items in the Bored context and the sum in the Interested context; therefore, positive SAI scores are indicative of more normative social adaptation, reflecting more questions, topic changes, and involvement, but less rapport when adapting to the Bored condition.

Social Interaction Anxiety Scale (SIAS). The SIAS is a 20-item self-report measure of general social interaction fears (e.g., speaking with an authority figure, being ignored; Mattick & Clarke, 1998). All items are rated on a 5-point scale based

on the degree to which respondents feel that the given statement is characteristic of them, so that higher SIAS total scores indicate greater levels of social interaction anxiety. The SIAS has demonstrated good internal consistency (α = .90), test–retest reliability (.92), convergent validity, and discriminant validity (Mattick & Clarke, 1998; Osman, Gutierrez, Barrios, Kopper, & Chiros, 1998). The SIAS has also shown sensitivity to post-treatment change (e.g., McEvoy, Nathan, Rapee, & Campbell, 2012).

Social Phobia and Anxiety Inventory–23 (SPAI-23). The SPAI-23 is a validated diagnostic screening instrument for social phobia in adults (Roberson-Nay, Strong, Nay, Beidel, & Turner, 2007). With only 23 items, it is an abbreviated version of the original SPAI (Turner, Beidel, Dancu, & Stanley, 1989). The SPAI-23 has demonstrated excellent internal consistency (α = .95; Roberson-Nay et al., 2007) and good test–retest stability (r = .72; Schry, Roberson-Nay, & White, 2012). All items are rated on a 5-point scale, ranging from *never* to *always*, with higher scores indicating more social anxiety. The SPAI-23 has a Social Phobia subscale (16 items) and an Agoraphobia subscale (7 items). The Agoraphobia subscale score is subtracted from the Social Phobia subscale score to compute the Difference score, which is considered the purest measure of social anxiety (i.e., with social avoidance related to agoraphobia removed).

Social Responsiveness Scale, Second Edition, Adult (SRS-2-A) and Adult and Self-Report (SRS-2-A-SR). The second edition of the SRS (SRS-2; Constantino & Gruber, 2012) includes a modification for adults of a validated scale of ASD severity in children (Constantino & Gruber, 2005). It is a 65-item rating scale that measures ASD-related social impairments, including social awareness, social information processing, reciprocal social communication, social motivation, and restricted interests/repetitive behaviors. The SRS-2 provides a *T*-score about the degree of interference in everyday life situations. The SRS-2 has a *T*-score range of 59 or less (normal range), 60 to 65 (mild range), 66 to 75 (moderate range), and 76 or greater (severe range). The SRS-2-A-SR was completed by the young adult with ASD (self-report), and the SRS-2-A (other-report) was completed by the young adult's assistant for the program.

Procedures

After ensuring eligibility of all five participants, pre-treatment assessments were completed 1 week prior to commencing the PEERS program. Immediately after the final group meeting, the post-treatment assessment occurred. All assessments took place at the clinic where the intervention was conducted. Pre-treatment and post-treatment CASS administrations were videotaped. Videotapes were marked so that, subsequently, the evaluators masked to assessment condition could not determine whether they were coding a pre- or post-treatment assessment, and all coding took place after study completion. The videotaped CASS administrations were coded by two trained raters independently, naïve to condition (pre- or post-treatment), after treatment completion.

Analyses

We compared the present sample's CASS scores with those of the Ratto et al. (2011) ASD and non-ASD samples, to estimate reliability of the measure across sites and samples, and examined interrater reliability in the present study. Agreement between raters occurred when they gave the same rating to a CASS domain or were within 1 point of each other (i.e., any discrepancy of 2 points or more was a disagreement). Interrater reliability was calculated as ([No. of Agreements/Total No. of Codes] × 100) for all CASS domains for all participants. Non-parametric correlations between pre-treatment I-CASS Total Score and pre-treatment scores from the SRS-2 (SRS-2-A and SRS-2-A-SR) were used to assess multi-method construct validity. Change scores were computed by subtracting pre-treatment scores from post-treatment scores for all measures. CASS scores were expected to be higher after treatment, indicating more socially appropriate behavior. We first examined change (pre- to post-treatment) in social behaviors within the Interested context only (I-CASS Scores) across the nine domains coded individually. We expected change (from pre- to post-treatment) in every CASS domain, except Gestures and Kinesic Arousal as there were no group (ASD, non-ASD) or interaction (context by group) effects for these two domains in the Ratto et al. study. For every other domain in the Ratto et al. study, the ASD group obtained lower scores (with Rapport reverse-scored). We also evaluated the degree to which the participants altered their behavior, or changed it when interacting with a conversational partner who appeared bored (B-CASS), by examining change in the SAI. Reliable change indices (RCI; Jacobson & Truax, 1991) were computed to determine whether change in SAI and I-CASS domain scores from pre-treatment to post-treatment was significant for each of the participants. RCI values greater than 1.96 indicate statistically significant and clinically meaningful change. We then explored the degree to which change on specific I-CASS items was related to change on indices of social behavior and social anxiety through non-parametric correlations.

Results

Reliability and Validity

Interrater reliability was high between two independent raters who coded all CASS administrations for the present study (88.89% agreement). The pre-treatment CASS

Table 3. RCI Post–Pre Treatment for the Core I-CASS Domains.

Participant	Asking questions	Topic changes	Overall involvement	Quality of rapport	Social anxiety
James	−0.569	1.042	5.770[a]	1.444	1.216
Mark	1.709	−3.126[a]	2.885[a]	1.444	1.216
Drew	4.558[a]	3.126[a]	8.656[a]	0.000	1.216
Anthony	−1.139	0.000	0.000	2.888[a]	1.216
Arnold	1.139	4.169[a]	2.885[a]	0.000	2.432[a]

Note. RCI = reliable change indices; I-CASS: Interested condition–Contextual Assessment of Social Skills.
[a]RCI > 1.96.

Table 4. Performance on the Interested and Bored Context CASS by Time Points.

Domain	Pre-treatment ($n = 5$)		Endpoint ($n = 5$)	
	Interested	Bored	Interested	Bored
Asking questions	4.80 (2.77)	5.80 (3.96)	6.80 (3.27)	8.00 (3.93)
Topic changes	1.60 (1.51)	3.20 (2.38)	2.60 (2.07)	4.20 (1.78)
Overall involvement	4.40 (0.54)	4.80 (1.09)	5.80 (1.09)	5.60 (0.894)
Quality of rapport	4.80 (1.09)	3.80 (1.78)	5.60 (0.54)	5.00 (0.71)
Kinesic arousal	5.60 (1.51)	5.00 (1.22)	5.60 (1.51)	5.00 (1.87)
Social anxiety	4.20 (1.30)	3.80 (1.92)	5.40 (1.34)	4.60 (1.14)
Vocal express	4.80 (1.30)	5.00 (1.22)	5.20 (1.48)	5.40 (1.34)
Gestures	4.40 (2.19)	4.20 (2.49)	5.00 (1.73)	4.80 (1.64)
Positive affect	5.40 (0.89)	4.20 (2.16)	5.80 (1.09)	4.80 (0.83)

Note. Raw scores reported as means and standard deviations. Asking questions and topic changes scored as behavior counts. All other variables scored on 1 to 7 scale, 1 = *low* to 7 = *high*. CASS = Contextual Assessment of Social Skills.

domain scores for the participants in the present study were similar to those obtained by Ratto and colleagues (2011) with a sample of 20 adults with ASD. Specifically, no differences between the current sample at pre-treatment and the Ratto et al. ASD sample were found (all I-CASS item t tests range from 0.067 to 1.489, *ns*), except for the Kinesic Arousal domain, $t(23) = 2.655$, $p = .014$, for which the present sample's mean (5.60 ± 1.51) was higher than that of the Ratto et al. sample (4.13 ± 1.00). There were also no significant group differences for domains on the B-CASS (t tests range from 0.081 to 1.949, all *ns*). Correlations between the pre-treatment I-CASS Total Score and the rating scales of social function were not significant, as the SRS-2-A-SR ($r = −.718$, $p = .172$) and SRS-2-A ($r = −.300$, $p = .624$) before treatment revealed no significant relationships.
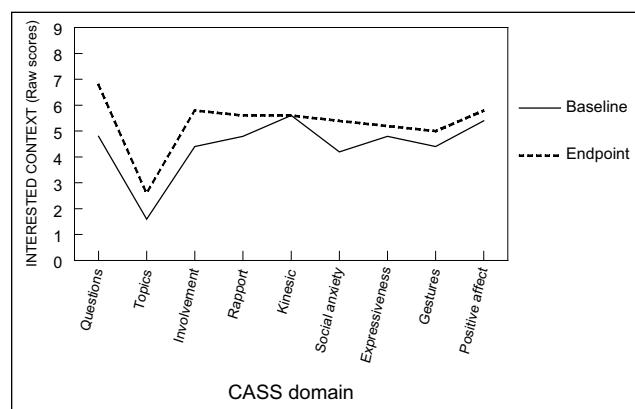
Sensitivity to Change

Table 3 shows the RCI values for the core areas of the I-CASS. Only Mark showed a significant increase on Asking Questions. Drew and Arnold demonstrated significant improvement in Topic Changes. Mark showed a decline on this item, although he performed significantly better at post-treatment for Overall Involvement. In fact, significant improvements were made by James, Mark, Drew, and Arnold in Overall Involvement, and Anthony performed significantly better on Overall Quality of Rapport at the end of the program. RCI values were also calculated for the exploratory domains on the CASS. Only James showed significant improvement in Gestures (RCI = 2.886) and Mark in Vocal Expressiveness (RCI = 3.637).

On the post-treatment SAI, James, Drew, and Anthony obtained positive z scores, indicative of more normative social adaptation (a positive difference from Bored to Interested context), ranging from 0.540 to 1.177, whereas before treatment, only Drew obtained a positive SAI score (SAI = 4.007). However, only James demonstrated statistically significant improvement on the SAI (RCI = 2.062) from pre-treatment to post-treatment. Observed change on the individual raw scores for the I-CASS and B-CASS domains was variable (see Table 4).

Figure 1. CASS (Interested condition) domain mean-raw scores at pre-treatment and post-treatment.
Note. CASS = Contextual Assessment of Social Skills.

Change as rated on specific I-CASS domains was related to change on other measures, in the expected direction. A significant correlation was found between change in Rapport and the SPAI-23 Difference score ($r = −.894$, $p = .041$). Change on the Involvement domain correlated significantly with change on the SIAS ($r = −.975$, $p = .005$). Likewise, change on the Gestures domain correlated with change on the SRS-2-A ($r = −.894$, $p = .041$). Changes on the SAI and I-CASS Total score were not significantly correlated with change on rating scales of either social function or anxiety. Despite the variability with respect to statistically significant change post-treatment on the CASS, as can be seen in Figures 1 and 2, across both conditions, the participants consistently showed improved social behavior (raw scores) after treatment in comparison with pre-treatment.

Discussion

The purpose of this study was to assess the utility, including validity and sensitivity to change, of the CASS as an
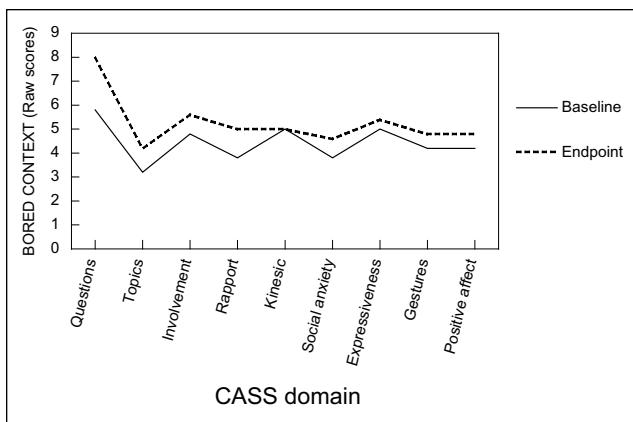
Figure 2. CASS (Bored condition) domain mean-raw scores at pre-treatment and post-treatment.
Note. CASS = Contextual Assessment of Social Skills.

outcome measure in treatment research with adults who have HFASD. Results indicate that the CASS could be implemented as intended by the developers in a treatment outcome study. Confederates were able to be trained efficiently, and raters were trained to reliably code behavioral observations during the CASS. Similarity in scores obtained by two samples of adults with ASD at two different sites, coded by different raters, supports cross-site and cross-sample reliability of the measure. Contrary to what was expected, though consistent with Ratto et al.'s (2011) findings, we did not find evidence that the CASS overlaps with other indicators of ASD severity (i.e., the SRS-2-A-SR, SRS-2-A). Although not statistically significant, this may be due to the small sample, as the magnitude of the relationship across measures was moderate to large.

The CASS may be a sensitive gauge of treatment response, at least with respect to specific social behaviors. Conceptualizing the Interested condition of the CASS as a better representation of typical social interactions, overall change (measured by the RCI) was not seen for all participants, nor was change consistent across all domains of social behavior. As a group, participants obtained higher (better) scores for most of the interested conversation domains at the end of the program compared with pre-treatment. The CASS domain that demonstrated the largest and most consistent improvement in this sample was Overall Involvement in the conversation, on which four participants showed significant improvement. This finding could have implications for use of the CASS as an outcome measure for samples with ASD. The definition of Overall Involvement given by Ratto and colleagues (2011) is indicative of its importance: "degree to which the participant's verbal and nonverbal behaviors indicated interest in the conversation and the proportion of the conversational burden carried by the participant" (p. 1281). Ratings of Overall Involvement were based on both verbal and non-verbal

indicators of interest or non-interest in the interaction "as a whole" (per the CASS rating manual, p. 6). This domain might capture meaningful social improvement and perhaps be less affected by factors outside the participant, such as the specific content of the conversation, something that can easily happen for some other CASS domains (e.g., Asking Questions and Topic Changes). Further study must be undertaken to determine whether the inconsistent improvement seen in the present study, across CASS domains and participants, reflects lack of uniformity in the measure's ability to detect change or something about the treatment (e.g., differential skill improvement across domains).

A small sample is the most prominent limitation of this study, limiting the conclusions that can be drawn. This study represents only a preliminary step toward validation of the CASS and, as such, must be followed by more robust multi-trait–multi-method approaches using larger samples. Notably, the Social Adaptation Index (which uses *z* scores) results should be interpreted carefully because of dependence on the sample characteristics. Ideally, we would also have a no-treatment comparison condition against which to assess stability of CASS scores over time. Finally, although participants had pre-existing clinical diagnoses of ASD, all of which were supported by SRS-2 data, diagnostic confirmation with either the *Autism Diagnostic Interview–Revised* (ADI-R; Le Counter, Lord, & Rutter, 2003) or *Autism Diagnostic Observation Schedule, Second Edition* (ADOS-2; Lord et al., 2012) would have been preferred. Future research on the utility of the CASS should use larger samples and evaluate skill generalization via assessment in more naturalistic settings.

These limitations notwithstanding, this is the first study to explore utility of a behavioral observational tool as an outcome measure with adults who have ASD. As such, we would like to summarize some aspects of the measure that should be considered in subsequent treatment research using the CASS or similar observational measures with adults. Although the confederates were trained and monitored (via live or video supervision) during the trial to ensure the conditions were delivered as intended, some differences across confederates were apparent during coding (e.g., some were more believable in the Bored context). Procedures to maintain consistency in confederates' social-communicative behavior during the CASS may help ensure reliability across participants and time points. Suggestions to improve consistency across administrations and to ensure the conversation is truly social and reciprocal in nature include limiting conversation about any restricted interests by having the confederate insert topic changes when such an interest arises (although this would require that the confederate be informed about special interests in advance) and modifying the instructions given at the start of the task (e.g., "You will have 3 minutes to talk and get to know each other, after which time you will be asked a few questions about the other person") to promote a reciprocal exchange and avoid discussion just on "things" or singular

events. Regardless of possible adaptations to the measure, the CASS possesses important characteristics necessary for sensitive and adequate observational tasks (Merrell, 1999), including well-defined observational domains, established procedures to determine interrater reliability, and an approach that makes discrete and unobtrusive observations possible, based on videotape review.

The CASS might be an especially useful supplemental outcome measure for gauging change in specific social behaviors, as well as for setting specific goals for a comprehensive approach to personalized intervention. A multi-dimensional approach needs to be taken when assessing changes in social functioning post-treatment. Further research is needed to assess the stability of behaviors coded with the CASS without treatment, as well as the measure's validity and sensitivity to improved social behavior as a function of treatment. Results from this pilot study, however, provide preliminary support for the CASS as part of a comprehensive, multi-modal assessment battery.

Acknowledgment

Authors' Note

During the conduct of this study, Saray Bonete was affiliated with the University of Granada. She is now with Francisco Vitoria University.

Declaration of Conflicting Interests

Funding

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Bellini, S., Peters, J. K., Benner, L., & Hopf, A. (2007). A meta-analysis of school-based social skills interventions for children with autism spectrum disorders. *Remedial and Special Education*, *28*, 153–162. doi:10.1177/07419325070280030401

Billstedt, E., Gillberg, C., & Gillberg, C. (2005). Autism after adolescence: Population-based 13- to 22-year follow-up study of 120 individuals with autism diagnosed in childhood. *Journal of Autism and Developmental Disorders*, *35*, 351–360. doi:10.1007/s10803-005-3302-5

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

Constantino, J. N., & Gruber, C. P. (2012). *Social Responsiveness Scale–Second edition* (SRS-2). Torrance, CA: Western Psychological Services.

Cunningham, A. B. (2011). Measuring change is social interaction skills of young children with autism. *Journal of Autism and Developmental Disorders*, *42*, 593–605. doi:10.1007/s10803-011-1280-3

Eaves, L. C., & Ho, H. L. (2008). Young adult outcome of autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*, 739–747. doi:10.1007/s10803-007-04410x

Gantman, A., Kapp, S. K., Orenski, K., & Laugeson, E. A. (2012). Social skills training for young adults with high-functioning autism spectrum disorders: A randomized controlled pilot study. *Journal of Autism and Developmental Disorders*, *42*, 1094–1103. doi:10.1007/s10803-011-1350-6

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System (SSRS)*. Circle Pines, MN: American Guidance Service.

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System–Rating Scales manual*. Minneapolis, MN: Pearson Assessments.

Hauck, M., Fein, D., Waterhouse, L., & Feinstein, C. (1995). Social initiations by autistic children to adults and other children. *Journal of Autism and Developmental Disorders*, *25*, 579–595. doi:10.1007/BF02178189

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychological research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19. doi:10.1037/0022-006x.59.1.12

Kasari, C., Rotheram-Fuller, E., & Locke, J. (2005). *The development of the playground observation of peer engagement (POPE) measure*. Unpublished manuscript, University of California Los Angeles.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test–Second edition*. Circle Pines, MN: American Guidance Service.

Kreiser, N. L., & White, S. W. (2014). Assessment of social anxiety in adolescents and adults with autism spectrum disorder. *Clinical Psychology: Science and Practice*, *21*, 18–31.

Laugeson, E. A., & Frankel, F. (2010). *Social skills for teenagers with developmental and autism spectrum disorders: The PEERS treatment manual*. New York, NY: Routledge.

Laugeson, E. A., & Frankel, F. (2014). *The PEERS treatment manual for young adults with autism spectrum disorders: Caregiver-assisted social skills training*. Manuscript in preparation.

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

Le Counter, A., Lord, C., & Rutter, M. (2003). *The Autism Diagnostic Interview–Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services.

Le Poire, B. A., & Yoshimura, S. M. (1999). The effects of expectancies and actual communication on nonverbal adaptation and communication outcomes: A test of interaction adaptation theory. *Communication Monographs*, *66*, 1–30. doi:10.1080/036377599093764660

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2002). *Autism Diagnostic Observation Schedule*. Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation*

Schedule–Second edition. Los Angeles, CA: Western Psychological Services.

Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., . . .Yoder, P. (2005). Challenges in evaluating psychosocial interventions for autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*, 695–708. doi:10.1007/ s10803-005-0017-6

Losh, M., Adolphs, R., Poe, M. D., Couture, S., Penn, D., Baranek, G. T., & Piven, J. (2009). Neuropsychological profile of autism and the broad autism phenotype. *Archives of General Psychiatry*, *66*, 518–526. doi:10.1001/archgenpsychiatry.2009.34

Luteijn, E. F., Jackson, S. A. E., Volkmar, F. R., & Minderaa, R. B. (1998). Brief report: The development of the children's social behavior questionnaire: Preliminary data. *Journal of Autism and Developmental Disorders*, *28*(6), 559–565.

Matson, J. L., Rotatori, A. F., & Helsel, W. J. (1983). Development of a rating scale to measure social skills in children: The Matson evaluation of social skills with youngsters (MESSY). *Behaviour Research and Therapy*, *21*, 335–340.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*, 455–470. doi:10.1016/S0005-7967(97)10031-6

Mazefsky, C. A., & White, S. W. (2013). Adults with autism. In S. J. Rogers (Ed.), *Handbook of autism spectrum disorders* (4th ed., pp. 191–211). Hoboken, NJ: Wiley.

McEvoy, P. M., Nathan, P., Rapee, R. M., & Campbell, B. N. C. (2012). Cognitive behavioural group therapy for social phobia: Evidence of transportability to community clinics. *Behaviour Research and Therapy*, *50*, 258–265. doi:10.1016/j.brat.2012.01.009

Merrell, K. W. (1999). *Behavioral, social and emotional assessment of children and adolescents*. Mahwah, NJ: Lawrence Erlbaum.

Merrell, K. W. (2001). Assessment of children's social skills: Recent developments, best practices, and new directions. *Exceptionality*, *9*, 3–18. doi:10.1080/09362835.2001.9666988

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Siebert, J. (2003). *A manual for the abridged Early Social Communication Scales (ESCS)*. Coral Gables, FL: University of Miami.

Osman, A., Gutierrez, P. M., Barrios, F. X., Kopper, B. A., & Chiros, C. E. (1998). The Social Phobia and Social Interaction Anxiety Scales: Evaluation of psychometric properties. *Journal of Psychopathology and Behavioral Assessment*, *20*, 249–264. doi:10.1023/A:1023067302227

Paul, R., Orlovski, S. M., Marcinko, H. C., & Volkmar, F. (2009). Conversational behaviors in youth with high-functioning ASD and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *39*, 115–125. doi:10.1007/s10803-008-0607-1

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

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*, 1277–1286. doi:10.1007/ s10803-010-1147-z

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

Roberson-Nay, R., Strong, D. R., Nay, W. T., Beidel, D. C., & Turner, S. M. (2007). Development of an abbreviated Social Phobia and Anxiety Inventory (SPAI) using item response theory: The SPAI-23. *Psychological Assessment*, *19*, 133–145. doi:10.1037/ 1040-3590.19.1.133

Schry, A. R., Roberson-Nay, R., & White, S. W. (2012). Measuring social anxiety in college students: A comprehensive evaluation of the psychometric properties of the SPAI-23. *Psychological Assessment*, *24*, 846–854. doi:10.1037/a0027398

Segrin, C., & Givertz, M. (2003). Methods of social skills training and development. In J. O. Greene & B. R. Burleson (Eds.), *Handbook of communication and social interaction skills* (pp. 135–178). Mahwah, NJ: Lawrence Erlbaum.

Smith, L. E., Maenner, M. J., & Seltzer, M. M. (2012). Developmental trajectories in adolescents and adults with autism: The case of daily living skills. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*, 622–631. doi:10.1016/j.jaac.2012.03.001

Turner, S. M., Beidel, D. C., Dancu, C. V., & Stanley, M. A. (1989). An empirically derived inventory to measure social fears and anxiety: The Social Phobia and Anxiety Inventory. *Psychological Assessment*, *1*, 35–40. doi:10.1037/1040-3590.1.1.35

Wang, P., & Spillane, A. (2009). Evidence-based social skills interventions for children with autism: A meta-analysis. *Education and Training in Developmental Disabilities*, *44*, 318–342.

White, S. W., Koenig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism and Developmental Disorders*, *37*, 1858–1868. doi:10.1007/ s10567-009-0062-3

White, S. W., & Maddox, B. B. (2013). Social interventions. In F. R. Volkmar (Ed.), *Encyclopedia of autism spectrum disorders* (Vol. 3, pp. 2896–2903). Berlin, Germany: Springer. doi:10.1007/SpringerReference_3342222012-12-1214:35:13 UTC

White, S. W., Ollendick, T., Albano, A., Oswald, D., Johnson, C., Southam-Gerow, M. A., . . .Scahill, L. (2013). Randomized controlled trial: Multimodal anxiety and social skill intervention for adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *43*, 382–294. doi:10.1007/s10803-012-1577-x

Wolery, M., & Garfinkle, A. N. (2002). Measures in intervention research with young children who have autism. *Journal of Autism and Developmental Disorders*, *32*, 463–478. doi:10. 1023/A:1020598023809