

Development of the Contextual Assessment of Social Skills (CASS): A Role Play Measure of Social Skill for Individuals with High-Functioning Autism

Allison B. Ratto · Lauren Turner-Brown ·

Betty M. Rupp · Gary B. Mesibov ·

David L. Penn

© Springer Science+Business Media, LLC 2010

Abstract This study piloted a role play assessment of conversational skills for adolescents and young adults with high-functioning autism/Asperger syndrome (HFA/AS). Participants completed two semi-structured role plays, in which social context was manipulated by changing the confederate's level of interest in the conversation. Participants' social behavior was rated via a behavioral coding system, and performance was compared across contexts and groups. An interaction effect was found for several items, whereby control participants showed significant change across context, while participants with HFA/AS showed little or no change. Total change across contexts was significantly correlated with related social constructs and significantly predicted ASD. The findings are discussed in terms of the potential utility of the CASS in the evaluation of social skill.

Keywords Autism spectrum disorders · Social skills · Adolescent

These results were presented in part at the International Meeting for Autism Research in Philadelphia, PA, May 2010.

A. B. Ratto · B. M. Rupp · D. L. Penn (✉)
Department of Psychology, University of North Carolina-Chapel Hill, Davie Hall, CB#3270, Chapel Hill, NC 27599-3270, USA
e-mail: dpenn@email.unc.edu

A. B. Ratto
e-mail: bassetta@email.unc.edu

L. Turner-Brown · D. L. Penn
Neurodevelopmental Disorders Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

G. B. Mesibov
Department of Psychiatry, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA

Introduction

Although the relative severity of autistic symptoms varies among individuals on the autism spectrum, deficits in social functioning are universal across the spectrum and generally persist throughout the life span (Mesibov et al. 2007). Social deficits can be especially problematic for those with high-functioning autism (HFA) or Asperger syndrome (AS) because these individuals are more aware of their social peculiarities, and thus are more negatively impacted by them (Tse et al. 2007). Social skills are especially critical in adolescence and young adulthood, as individuals with HFA/AS move toward independent lives. As awareness of the social needs of adolescents and young adults with HFA/AS has increased, research on social skills interventions for this population has increased as well (Tse et al. 2007). The research on social skills assessment for this population, however, has not advanced at the same pace (Matson and Wilkins 2007). The most commonly used method for assessing social skills intervention outcomes in individuals with HFA/AS remains informant report, usually in the form of questionnaires (Matson and Wilkins 2007). While questionnaire methods are quick and cost effective, their results are limited by their dependence on the informant's perspective and biases (Bellack et al. 2006). Many studies address this limitation by combining third-party questionnaires with laboratory-based measures that target discrete areas of social functioning (e.g. theory of mind) (Webb et al. 2004). Unfortunately, improvements on these measures often do not generalize to daily social functioning (White et al. 2007).

In order to better evaluate treatment generalization, some studies have employed observational assessments of younger children with ASD, usually focusing on peer interactions (Matson and Wilkins 2007; White et al. 2007).

In research with adolescents and young adults, however, direct observational assessment is more challenging, as they are less likely than young children to be in easily observable social settings (e.g. recess, “circle time”) (Matson and Wilkins 2007). One form of observational assessment that can be readily used with adolescents and adults is a role play task. While a few intervention studies for adolescents and adults with ASD have used role plays as an informal outcome measure (Paul 2003; Webb et al. 2004), a systematic role play assessment has not yet been developed for individuals with HFA/AS. Furthermore, while typical peers have been widely used in studies of young children with ASD, they are rarely utilized in the assessment of adolescents and adults (Matson and Wilkins 2007). This is a significant oversight because peer interactions are the primary social context for adolescents and young adults (Englund et al. 2000).

An important aspect of successful peer interactions is responding appropriately to nonverbal social cues, a particularly challenging skill for individuals with HFA/AS (Golan et al. 2006). Adolescents and adults with HFA/AS can use nonverbal cues to correctly interpret basic emotions such as happiness and sadness, but often have difficulty accurately perceiving more complex emotions such as boredom and resentment (Golan et al. 2006). Boredom is a particularly salient emotion for this population, as individuals with HFA/AS often do not realize when they are boring others by talking about their circumscribed interests (Mesibov et al. 2007). Typical individuals, by contrast, are more easily able to discern when a conversational partner is bored with the conversation and will alter their own behavior to keep the conversation going by asking questions or changing the topic of conversation (Burgoon et al. 1995; Kearsley 1976; McLaughlin and Cody 1982). Additionally, typical individuals generally show increased signs of social anxiety when speaking with a disinterested conversational partner, as indicated by a more tense posture and higher kinesic arousal (Burgoon et al. 1995; Williams and Zadro 2001). Thus, analyzing the behavioral response of adolescents and young adults with HFA/AS to a bored conversational partner would provide a more realistic behavioral measure of conversational skills, which greatly contribute to overall social skill, for this population.

The goal of this study was to evaluate a peer-enacted role play measure of conversational skill in which confederates displayed either social interest or boredom in two separate role plays. The first aim of the study was to establish reliability of this measure, via evaluation of inter-rater reliability and internal consistency. The second aim was to evaluate participants’ ability to adapt their behavior in response to changes in social context. It was hypothesized that typical controls would ask more questions,

change the topic of conversation more frequently, increase their overall involvement, and show an increase in anxiety in the boredom role play relative to their behavior in the interested role play. In contrast, the HFA/AS group was expected to show stable levels of these behaviors across both role plays. In addition, it was hypothesized that ratings of conversational rapport would be stable across the two role plays for the HFA/AS group, but would be decreased for typical controls in the bored relative to the interested role play. The final aim was to establish convergent validity through calculating correlations of changes in role play behavior with verbal IQ, theory of mind, and autism severity, and to establish discriminant validity via correlations with performance IQ. In prior studies of individuals with ASD, verbal IQ, but not performance IQ, has been shown to correlate with measures of social skill (Golan et al. 2006). Predictive validity was examined in the ability of the CASS to predict ASD diagnosis.

Methods

Participants

Forty adolescents and young adults participated in the present study: 20 participants with high-functioning autism (HFA) or Asperger syndrome (AS), recruited from an intervention study providing Social Cognition Interaction Training for Autism (SCIT-A; Penn and Turner-Brown, PIs), and 20 control participants, recruited from an undergraduate population. All participants were required to be between the ages of 16 and 22 years, have a verbal IQ of 85 or higher, and speak English as a primary language. Additionally, participants in the HFA/AS group were required to have an existing clinical diagnosis and to meet criteria for ASD on the Autism Diagnostic Observation Schedule (ADOS). All HFA/AS participants were tested at baseline.

Measures

Screening Measures

All HFA/AS participants were administered a phone screening to determine eligibility, and typical participants completed a self-report screening form. Potential participants with a diagnosis of a disorder that may impair social functioning, other than HFA/AS (e.g. schizophrenia, social phobia) were excluded from the study. In addition to the general screening form, typical controls also completed the Baron-Cohen Autism Spectrum Quotient (AQ; Baron-Cohen et al. 2001). The AQ is a 50-item self-report measure of autism spectrum symptoms normed for adults ages

Table 1 Demographic variable by group

	Control	HFA/AS
N	20	20
Gender	85% Male (<i>n</i> = 17)	85% Male (<i>n</i> = 17)
Ethnicity	White: 90% (<i>n</i> = 18) Latino: 5% (<i>n</i> = 1) Asian: 5% (<i>n</i> = 1)	White: 90% (<i>n</i> = 18) Latino: 5% (<i>n</i> = 1) Asian: 5% (<i>n</i> = 1)
Age	19.9 years (1.2)	17.9 years (1.4)**
Full scale IQ	113.6 (8.3)	108.1 (14.8)
Verbal IQ	115.8 (9.8)	109.1 (16.3)
Performance IQ	108.3 (8.8)	105.7 (14.4)
Theory of Mind	52.6 (4.0)	46.0 (6.7)*
Autism Quotient	12.9 (3.0)	—

Entries in the table are means and standard deviations. Dashes indicate that data was not collected

HFA/AS, High-functioning autism/Asperger syndrome; IQ scores reported as standard scores. Theory of mind, raw score on The Awareness of Social Inference Test (TASIT); Autism Quotient, raw score on the Baron-Cohen Autism Spectrum Quotient (AQ)

* *p* < .0005, ** *p* < .0001

16 and up. Scores range from 0 to 50, and the established cut-off of 32 was used to exclude potential participants from the control group (Table 1). Using the lower cut-off of 26, as reported in some studies, would not have eliminated any participants from the control group (Woodbury-Smith et al. 2005).

Cognitive Ability Measure

All participants were administered the Wechsler Abbreviated Scales of Intelligence by a trained research assistant (WASI; Wechsler 1999).

Theory of Mind

The Awareness of Social Inference Test (TASIT; McDonald et al. 2003) is a three-part test of social perception and theory of mind. Subtest two, *Social Inference-Minimal*, used in the present study, assesses theory of mind skills by asking individuals to interpret the meaning and intentions behind potentially ambiguous remarks made by actors in video vignettes. One point is given for each correct response and then summed to calculate a norm-referenced total score. The TASIT has demonstrated convergent validity with several measures of social functioning and theory of mind and has adequate test-retest reliability (McDonald et al. 2006). Within the present study, the TASIT also demonstrated high internal consistency ($\alpha = .71$).

Parent Report of Autism Severity

The Social Responsiveness Scale (SRS; Constantino et al. 2000) is a 65-item questionnaire that asks the informant to rate the individual's level of reciprocal social behavior and autism symptoms. The SRS generates a total t-score that denotes the individual's degree of social impairment from autism symptoms, where higher scores indicate more severe impairment. The scale has demonstrated high construct and discriminant validity and high test-retest reliability in prior research (Constantino et al. 2000; Constantino and Todd 2003). Parents completed the SRS for participants in the HFA/AS group. The SRS was not administered to the control group.

Target Measure: The Contextual Assessment of Social Skills (CASS)

Procedure

In the CASS, participants have two role play conversations with two different confederates, each of which lasts 3 min. Participants speak only to confederates of the opposite gender, as one of the primary social tasks of the adolescent/young adult period is to master opposite gender interactions (Connolly et al. 2004; Paul 2003). Prior to each conversation, the examiner reads the following prompt to the participant and the confederate:

Thank you both so much for coming in. Right now we'd like for each of you to act as if you had recently joined a new club or social group, and now you're sitting next to each other, waiting for the first meeting of this new club or group to start. You will have 3 min to talk to each other, and then I will come back in the room.

The examiner then exits the room. After 3 min, the examiner re-enters the room, escorts the confederate out, and asks the participant to complete a brief questionnaire about the conversation with the confederate (*Conversation Rating Scale*, described below).

Though the set-up of the two role plays is identical, the behavior of the confederates differs markedly for each, creating two distinct social contexts. In the first role play, the confederate demonstrates social interest and engagement in the conversation, and in the second, the confederate indicates boredom and disengagement. We originally planned to counter-balance the order of social context. However, pilot testing and further literature review led to the decision to have participants experience the interested context before the bored context. Expectations about another's behavior guide social interactions, and violations of those expectations produce changes in behavior as

individuals adapt to the social situation and respond to their partner's deviation from the norm (LePoire and Yoshimura 1999). The confederate's behavior in the interested context represents expected, typical behavior, while the bored context represents a violation of social expectations (LePoire and Yoshimura 1999). Presenting the interested context first allows for rating of the participant's social behavior in a normative interaction (the interested context), and the participant's adaptation to a social change (the bored context). Presenting the bored context first would violate social expectations and lead the participant to expect similar behavior in the second (interested) context, thus leading to an inaccurate representation of the participant's behavior in a normative social interaction (Burgoon and Hale 1987; LePoire and Yoshimura 1999).

In each context, the confederate's eye contact, facial affect, posture, and gestures are manipulated to create either the interested or the bored context. In addition, confederates ask questions and elaborate on statements in the interested context, but minimize social initiation and keep statements brief in the bored context. Across both role plays, it is important that the confederate not carry the conversational burden, in order to allow for an accurate assessment of the participant's social skill. Consequently, confederates are instructed to speak for no more than 50% of the time and to wait 10 s after the examiner leaves the room for the participant to initiate the conversation before starting the conversation themselves. Standard prompts are provided for initiation when necessary and are identical across role plays. If conversational lapses occur, confederates wait 5 s before reinitiating the conversation in the interested context, and 7 s in the bored context. Conversational pauses longer than 3 s are perceived as awkward by typical individuals, and thus are quickly filled (Tree 2002; McLaughlin and Cody 1982). The extended wait times described above are used in the present study to allow for the slower social processing time of individuals with ASD, and to keep the conversational burden on the participant. The slightly longer wait time is used in the bored context to maintain the differences between the bored and interested contexts (Tree 2002). Confederates were 9 undergraduate students, ages 18–21, who completed 2 h of training prior to participation in the study and were randomly assigned to role plays using the Urn Randomization Program, designed for balanced randomization of groups in small samples. The lead author watched the role play videotapes and provided systematic feedback regarding fidelity to the confederates following each role play.

Behavioral Coding

All conversations were videotaped and participants' verbal and nonverbal behaviors were coded in ten categories:

Asking Questions, Topic Changes, Vocal Expressiveness, Gestures, Positive Affect, Posture, Kinesic Arousal, Social Anxiety, Overall Involvement in the Conversation, and Overall Quality of Rapport. These items were based on behavior which are indicative of conversational engagement and which are prone to change in response to conversational boredom (Burgoon et al. 1995). Several of these behaviors also overlap with coding items from the ADOS (Lord et al. 2000).

The four primary outcomes were Asking Questions, Topic Changes, Overall Involvement, and Overall Quality of Rapport, as these items have been previously found to change in response to perceived boredom and to be indicative of social skill. Social anxiety was also examined in this study through three rating items: a global social anxiety rating, posture, and kinesic arousal. Exploratory analyses were also conducted on vocal expressiveness, gestures, and positive affect, as these behaviors are often impaired in ASD but have not been shown to consistently change in response to perceived boredom. A description of each outcome rating can be found in Table 2.

Two trained raters coded each set of role plays. The items Asking Questions and Topic Changes were coded as behavioral counts; all other items were rated on a scale of 1–7 (1 = low, 7 = high). Raters were trained using a set of ten training videos. Reliability analyses were calculated using the last six videos rated, by partitioning the sum of squares for raters scores into error variance and true score variance, where true score variance was defined as the author's scores (Whitehurst 1984). Both raters achieved agreement with the lead author (ABR) within one interval rating of at least .70 in reliability analyses ($SS_{true}^2 = .73, .71$), before proceeding to rating study videos. Following training, raters' reliability was calculated with one another; the lead author did not rate any study videos. Raters were kept blind to study hypotheses and group membership.

Conversation Rating Scale

Following each role play conversation, participants completed the Conversation Rating Scale (CRS), a questionnaire developed for this study. The CRS asks the participant to rate the confederate's interest in the conversation, using five items rated on a 7-point Likert scale. The CRS items are based on items from the Interpersonal Communication Satisfaction Inventory (Hecht 1978) and the Relational Communication Scale (Burgoon and Hale 1987), two interpersonal communication rating scales that have been extensively validated in the communication literature (Graham 1994). The five items ask the participant to rate perceived interest, confederate friendliness, conversational flow, perceived boredom, and sense of distance. Item scores are summed to generate a total score of

Table 2 Description of outcomes of the contextual assessment of social skills (CASS)

Variable	Description
<i>Primary outcomes</i>	
Asking questions	The number of questions asked by the participant to engage the confederate in conversation
Topic changes	The number of times a participant used a question or comment to attempt to change the topic of conversation
Overall involvement	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
Overall quality of rapport	The level of comfort and the balance of conversational burden in the interaction
<i>Additional outcomes</i>	
Social anxiety (global)	Degree to which the participant's verbal and nonverbal behaviors were indicative of anxiety (e.g. fidgeting, vocal tremors, direct comments on anxiety)
Kinesic arousal	Intensity and frequency of body movement by the participant (e.g. foot tapping, leg shaking)
Posture	The degree of tension or relaxation in the participant's posture; <i>eliminated due to low variance</i>
Vocal expressiveness	Degree to which the participant varied his/her pitch, tone, and tempo
Gestures	Frequency and quality of participant's gestures, including conventional, descriptive, and emotional/emphatic gestures
Positive affect	Level of positive emotion indicated by the participant's facial expression and body language

perceived conversational interest for each social context. The total scores range from 5 to 35. Internal consistency for the CRS was high ($\alpha = .92$).

Results

All data analyses were performed using the SAS 9.1× system for Windows, and statistical significance was set at $p < .05$. Prior to any further analyses, the frequency distributions of each of the 10 behavioral items coded in the CASS were examined for the entire sample. The Posture item exhibited extremely low variance across contexts and groups, and thus was excluded from all subsequent analyses.

Sample Characteristics

Participants in the control group were matched to participants in the HFA/AS group on gender and ethnicity. The samples were 85% male and 90% Caucasian. There were no significant differences between the two groups on full scale IQ, verbal IQ, or performance IQ (Table 1). The control group was significantly older than the HFA/AS group ($t = 4.83$, $p < .0001$). As expected, the control group also had significantly higher scores on the theory of mind measure ($t = 3.81$, $p < .0005$).

Reliability

Inter-rater reliability was assessed by calculating the intraclass correlation coefficient (ICC) for each of the 9 items separately. ICC (3, 1) for a two-way mixed, random

effects model was used (Shrout and Fleiss 1979). The ICC values for all nine CASS items ranged from .50 to .97, with a mean value of .68. On the CASS primary outcomes (Asking Questions, Topic Changes, Overall Involvement and Overall Quality of Rapport), ICC values ranged from .62 to .96, with a mean ICC of .81. The mean ICC values of .68 for all items and .81 for items in the CASS total score were considered acceptable for further analyses. Each rater's scores were averaged to calculate a final score for each participant on each item, which were used for all subsequent analyses. Internal consistency of all 9 items on the CASS was high (standardized alpha = .83). Internal consistency was also analyzed separately for the four primary outcomes (Asking Questions, Topic Changes, Overall Involvement, and Overall Quality of Rapport) and was acceptable (standardized alpha = .75).

Manipulation Check

To verify that the two role play contexts were distinct from one another, the Conversation Rating Scale (CRS) was administered to participants following each role play. We expected a significant interaction effect, whereby the control participants would report significantly lower ratings on the CRS in the bored context (relative to the interested context), and the HFA/AS group should report little or no change. CRS total scores range from 5 to 35 (5 to 32 in the present sample), with higher scores indicating higher levels of perceived interest. A multiple linear regression analysis was used to evaluate the results of the CRS, representing autism status and context as dummy variables. A significant main effect was found for social context ($t = -8.35$, $p < .0001$), but not for autism status ($t = -.03$, ns). This

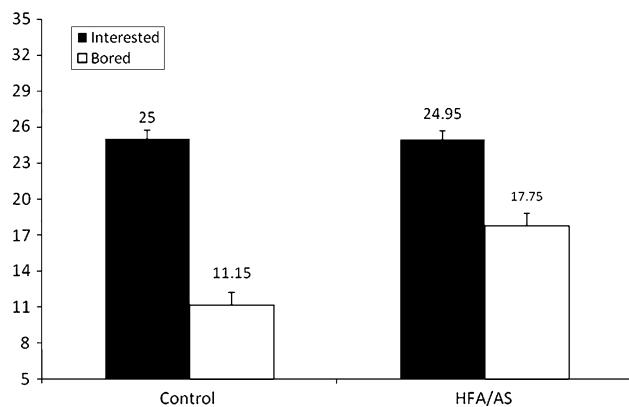


Fig. 1 CRS conversation rating scale, range = 5–35. HFA/AS = High-functioning autism/Asperger syndrome

indicated that across groups, ratings on the CRS were lower in the bored than in the interested context, but there were no differences overall by autism status. The interaction effect was also statistically significant ($t = 2.83, p < .006$), indicating that there was a significantly larger decrease in CRS ratings across context for the control group than for the HFA/AS group (Fig. 1).

This interaction was further probed using t-tests to check for group differences by context. There was no significant difference between the control and the HFA/AS group for the interested context ($t = .03, ns$), but for the bored context, the control group's ratings were statistically significantly lower than the HFA/AS group's ($t = -3.67, p < .001$).

CASS Group Differences (Primary Outcomes)

Mean scores by group and context for all individual CASS items can be found in Table 3.

For these analyses, dummy variables representing social context, autism diagnosis, and the interaction of these two factors were entered simultaneously into a linear regression model to predict score on each of the nine behavioral items of the CASS. Planned analyses were first undertaken on the primary outcomes: Asking Questions, Topic Changes, Overall Involvement, and Overall Quality of Rapport. For Asking Questions, the full model including context, autism diagnosis, and their interaction significantly predicted number of questions asked ($F = 16.12, p < .0001$) and accounted for a modest proportion of the variance in this item ($R^2 = .39$). The main effects for context and for autism diagnosis were both statistically significant ($t = 2.97, p < .01, t = -3.20, p < .01$). Across both groups, participants asked significantly more questions in the bored than in the interested context, and the control group asked significantly more questions of the confederate overall than did the HFA/AS group. The interaction effect for the model approached statistical significance ($t = -1.70, p < .09$), indicating that the control group showed a larger increase than the HFA/AS group in the number of questions asked in the bored compared to the interested context.

The full linear regression model for Topic Changes also significantly predicted scores on this item ($F = 14.38, p < .0001$) and accounted for a modest proportion of item variance ($R^2 = .36$). The main effects for both context ($t = 2.48, p < .02$) and autism diagnosis ($t = -3.36, p < .01$) were statistically significant, indicating that across groups, participants introduced topic changes more frequently in the bored than in the interested context, and that the control group introduced significantly more topic changes than did the HFA/AS group. The interaction of group and context was not statistically significant for this model ($t = -1.30, ns$), indicating that the increase in topic changes from the

Table 3 Performance on the CASS by group

	Control (n = 20)		HFA/AS (n = 20)	
	Interested	Bored	Interested	Bored
Asking questions	10.65 (4.07)	15.13 (5.01)	5.83 (4.83)	6.60 (5.09)
Topic changes	4.98 (2.20)	6.60 (2.30)	2.78 (1.82)	3.20 (1.93)
Overall involvement	6.08 (.44)	5.70 (.57)	4.93 (1.48)	4.90 (1.27)
Overall quality of rapport	6.13 (.53)	3.33 (.80)	4.88 (1.27)	3.58 (.91)
Social anxiety	5.58 (1.17)	4.80 (.92)	3.90 (1.45)	3.73 (1.32)
Kinesic arousal	4.28 (.87)	4.05 (.69)	4.13 (1.00)	3.95 (1.12)
Vocal expressiveness	5.80 (.57)	5.45 (.51)	4.55 (1.34)	4.65 (1.17)
Gestures	4.75 (1.24)	3.25 (1.58)	4.03 (2.14)	3.58 (2.14)
Positive affect	5.75 (.53)	5.03 (.47)	4.50 (1.81)	4.13 (1.60)
Posture	5.88 (.22)	5.83 (.29)	5.75 (.85)	5.65 (.75)

Scores reported as means and standard deviations. HFA/AS, high-functioning autism/Asperger syndrome. Asking questions and topic changes scored as behavior counts. All other variables scored on 1–7 scale, 1 = low, 7 = high. Posture not included in later analyses due to low variance

bored to the interested context did not differ significantly between the control group and the HFA/AS group.

Overall Involvement was also significantly predicted by the full model ($F = 6.29, p < .001$), though it accounted for a smaller proportion of item variance ($R^2 = .20$). The main effect for autism diagnosis was statistically significant ($t = -3.50, p < .001$), indicating that the control group was significantly more involved in the conversation than the HFA/AS group across social context. The main effect for context ($t = -1.14, ns$) and the interaction effect ($t = .75, ns$) were not statistically significant, indicating that overall involvement did not change significantly across contexts in either group.

The full linear regression model significantly predicted overall quality of rapport ($F = 39.92, p < .0001$) and accounted for a large proportion of the variance ($R^2 = .61$). The main effects for both context ($t = -9.68, p < .0001$) and autism diagnosis ($t = -4.32, p < .0001$) were statistically significant; across both groups, rapport was significantly lower in the bored relative to the interested context, and rapport was significantly lower in the HFA/AS group than in the control group in both contexts. The interaction effect was also statistically significant ($t = 3.67, p < .0001$), indicating that the control group showed a significantly larger decrease in quality of rapport in the bored compared to the interested context than the HFA/AS group.

The full model for social anxiety was also statistically significant ($F = 9.69, p < .0001$) and accounted for a modest proportion of the variance ($R^2 = .28$). The main effects for both context ($t = -1.99, p = .05$) and autism diagnosis ($t = -4.30, p < .0001$) were statistically significant. Across groups, social anxiety decreased in the bored relative to the interested context, and the HFA/AS group had significantly lower levels of social anxiety overall than the control group. The interaction effect of context and group was not statistically significant ($t = 1.09, ns$), indicating that the decrease in social anxiety across contexts was not significantly different between the two groups.

CASS Group Differences (Exploratory Analyses)

Exploratory analyses were also conducted to examine the main and interaction effects of group and context on Vocal Expressiveness, Gestures, Positive Affect, and Kinesic Arousal. Vocal Expressiveness was significantly predicted by autism diagnosis ($F = 22.61, p < .0001$), which accounted for a small proportion of the variance ($R^2 = .22$), with the HFA/AS group showing significantly lower vocal expressiveness overall than the control group. Adding context into the model did not have a significant effect on the R^2 value (R^2 increment = .0033, $F^2 = .004$), indicating that vocal expressiveness did not change

significantly across context for either group, and thus interaction effects were not tested. Gestures were best predicted by the full model ($F = 2.56, p = .06$), but the model only accounted for a very small proportion of the variance ($R^2 = .09$). Although the main effect for group was not statistically significant ($t = -1.26, ns$), the main effect for context was ($t = -2.61, p < .01$), indicating that both groups showed a decrease in gestures in the bored relative to the interested context. The interaction effect for this model was not statistically significant ($t = 1.29, ns$).

Both autism diagnosis and context emerged as significant predictors of Positive Affect ($F = 9.24, p < .001$), together accounting for a small proportion of the variance ($R^2 = .19$). The results indicated that the HFA/AS group showed significantly less positive affect overall than did the control group, and that across groups, participants showed a significant decrease in positive affect in the bored relative to the interested context. Adding the interaction effect into the model did not significantly increase the R^2 value (R^2 increment = .0041, $F^2 = .005$), so this effect was not interpreted. Kinesic Arousal was not significantly predicted by autism diagnosis ($F = .36, ns$) or by the combined model including context ($F = .65, ns$); thus the interaction effects were not examined. Finally, since the two groups significantly differed in age, this variable was entered as a covariate in all analyses; the results were unchanged.

Validity Analyses

The primary outcomes on the CASS (Asking Questions, Topic Changes, Overall Involvement, and Overall Quality of Rapport) were combined to create a CASS total score for the validity analyses. Normative social adaptation was represented by an increase in Asking Questions, Topic Changes, and Overall Involvement, but by a decrease in Overall Quality of Rapport across social contexts. Thus, the rapport item was reverse scored for these analyses only, so that an increase on all four items would correspond to normative social adaptation. Each of these four items were standardized (converted to z-scores) so that all would be on comparable scales. A total score was then calculated for each social context (interested and bored) by summing the standardized scores for all four items for each context. A CASS total change score was then calculated by subtracting the total score on the interested context from the total score on the bored context, so that higher change scores were indicative of more normative social adaptation.

CASS total change scores were modestly and significantly correlated with both verbal IQ ($r = .32, p < .04$) and theory of mind ($r = .47, p < .002$) across groups. Correlations were also conducted with autism severity, and contrary to expectations, this correlation was not

statistically significant ($r = -.22$, ns). In regard to discriminant validity, the CASS was not significantly associated with performance IQ, as predicted ($r = .006$, ns).

Predictive validity of the CASS would be supported if the total change score discriminates between individuals with HFA/AS and typical controls. The mean CASS change score (standardized score) for the control group was .91 ($SD = 2.03$), while the mean for the HFA/AS group was $- .91$ ($SD = 2.09$). A student's t-test indicated that the difference in means was statistically significant ($t = 2.80$, $p < .008$). A logistic regression model was also used to analyze the predictive validity, in which the CASS total change score was used to predict probability of an autism diagnosis. This model was statistically significant ($\chi^2 = 7.63$, $p < .006$), generating an odds ratio of .62. Thus, for every one unit increase in the CASS total change score, the probability of having an autism diagnosis was reduced by a factor of .62.

Discussion

The first aim of this study was to evaluate the reliability of the CASS via internal consistency and inter-rater reliability. Internal consistency was quite high, while inter-rater reliability was generally acceptable; the mean ICC value was near .70 and higher when only primary outcomes were considered. Given that ICC values were as low as .50 for some items, further development of the coding system to improve reliability will be an important goal in future work. The second aim of this study was to evaluate differences between typical controls and individuals with HFA/AS in the ability to adapt to changes in social context. As expected, typical controls accurately perceived changes in social context, rating confederates as significantly less interested in them in the bored relative to the interested context. Individuals with HFA/AS also perceived these changes in social context, though the difference in their ratings between contexts was less robust.

Consistent with prior research, the control group showed higher levels overall of asking questions, topic changes, involvement, and quality of rapport than the HFA/AS group. Additionally, across all participants, there was a significant increase in asking questions and topic changes, and a significant decrease in overall quality of rapport in the bored context. These results indicate that on these variables, the CASS does effectively discriminate between the two groups and detects differences in behavior by social context. Moreover, for asking questions, topic changes, and overall quality of rapport, there was an interaction effect whereby the control group generally demonstrated higher levels of change across context than did the HFA/AS group, which showed minimal to no

change. It should be noted that this interaction effect was only statistically significant for overall quality of rapport and approached statistical significance for asking questions. Thus, the pattern of the results was generally consistent with the study hypotheses, but subsequent research should be conducted with a larger sample to better discern interaction effects (Note: statistical power was approximately .50 for the linear regression analysis).

Contrary to expectations, no significant differences were observed in overall involvement across context in either group, and a slight decrease was observed in social anxiety. The lack of change in overall involvement may be accounted for by the brief time period of the role plays. Prior research on social adaptation has typically utilized longer interactions of 10–15 min, as opposed to the 3 min used in the present study, and found that overall involvement decreased slowly over the course of the interaction (Burgoon et al. 1995; LePoire and Yoshimura 1999). The shortened interaction time may not have allowed for the gradual changes in the global variable of involvement observed in prior research. Similarly, the slight decrease in social anxiety across contexts may also be explained by the brief time period used. Social anxiety is often initially high and decreases over time, particularly if participants are aware they are being videotaped, as in the present study (Burgoon et al. 1995; Williams and Zadro 2001). It is possible, then, that the decrease in social anxiety observed in the present study was due primarily to participants' adaptation to being videotaped, and is not a response to change in context.

The final aim of the study was to examine the construct validity of the CASS. In general, there was support for convergent and discriminant validity of the CASS, as the total change score of the CASS was associated with both verbal IQ and theory of mind, but was uncorrelated with performance IQ. The CASS total change score also significantly predicted the presence of ASD, supporting its predictive validity. Thus, the psychometric properties of the CASS appear sound and indicate that it has the potential to serve as a valid and reliable measure of social impairment in individuals with HFA/AS.

This study had a number of limitations. First, although our sample size is comparable to previous research that has examined social skills in HFA/AS (Golan et al. 2006; Tse et al. 2007), a larger sample is needed to extend the results found in the present study. Second, little is known about the stability of performance on the CASS over time or its sensitivity to treatment effects. Additionally, although the control sample was matched with the HFA/AS sample in ethnicity and IQ, it was a convenience sample of undergraduate students. Thus, these findings need to be replicated with a control sample drawn from the community at large. It may also be informative to examine the effects of extending

the time of the role plays, as well as the differences in interacting with same versus opposite-gender confederates.

In summary, this initial investigation of the CASS found that it has sound psychometric properties and was able to distinguish the conversational skill of individuals with HFA/AS from typical controls. For some key social behaviors, typical controls showed a pattern of adaptation to social context that was not observed in the HFA/AS group. Thus, the CASS has the potential to fulfill the need for an ecologically valid measure of social skill, which has been lacking in this field for quite some time (Lord et al. 2005). The CASS may also provide information about appropriate targets for treatment and about the impact of treatment on meaningful social behaviors. The current study underscores the promise of the CASS for studying social skill in HFA/AS, and the need to conduct future research on this instrument.

Acknowledgments This project was part of a larger study, an investigation of social cognition and interaction training for adolescents for autism (SCIT-A) funded by Autism Speaks (Penn, PI, Turner-Brown, Co-PI; Grant #1949). Renee Clark of the Research Participant Registry Core of the Carolina Institute for Developmental Disabilities, which is supported by the Eunice Shriver Kennedy National Institute of Child Health and Human Development, provided recruitment resources. Joseph Piven, Ph.D. is the director of the Carolina Institute for Developmental Disabilities, which provided the infrastructure for this research project. Chris Weisen, Ph.D. of the Odum Institute at UNC-Chapel Hill provided statistical consulting. The following UNC-Chapel Hill undergraduate students served as research assistants: Paige Allmendinger, Bianca Brooks, Sierra Carter, Lauren Catalano, Alexis Georges, Ben Gootman, Tiffany Hollung, Colin Iwanski, Caroline Oates, Anderson Shackleford, and Jeremy Weaver.

References

- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disabilities*, 31, 5–17. doi:10.1023/A:1005653411471.
- Bellack, A. S., Brown, C. H., & Thomas-Lohrman, S. (2006). Psychometric characteristics of role-play assessments of social skill in schizophrenia. *Behavior Therapy*, 37, 339–352. doi:10.1016/j.beth.2006.01.005.
- Burgoon, J. K., & Hale, J. L. (1987). Validation and measurement of the fundamental themes of relational communication. *Communication Monographs*, 54, 19–41.
- Burgoon, J. K., Stern, L. A., & Dillman, L. (1995). *Interpersonal adaptation: Dyadic interaction patterns*. New York, NY: Cambridge University Press.
- Connolly, J., Craig, W., Goldberg, A., & Pepler, D. (2004). Mixed-gender groups, dating, and romantic relationships in early adolescence. *Journal of Research on Adolescence*, 14, 185–207. doi:10.1111/j.1532-7795.2004.01402003.x.
- Constantino, J. N., & Todd, R. D. (2003). Autistic traits in the general population: A twin study. *Archives of General Psychiatry*, 60, 524–530. doi:10.1001/archpsyc.60.5.524.
- Constantino, J. N., Pryzbeck, T., Friesen, D., & Todd, R. D. (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *Journal of Developmental and Behavioral Pediatrics*, 21, 2–11.
- Englund, M. M., Levy, A. K., Hyson, D. M., & Sroufe, L. A. (2000). Adolescent social competence: Effectiveness in a group setting. *Child Development*, 71(4), 1049–1060.
- Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders*, 36(2), 169–183. doi:10.1007/s10803-005-0057-y.
- Graham, E. E. (1994). Relational communication scale. In R. B. Rubin, P. Palmgreen, & H. E. Sypher (Eds.), *Communication research measures: A sourcebook*. New York, NY: The Guilford Press.
- Hecht, M. L. (1978). Measures of communication satisfaction. *Human Communication Research*, 4, 350–368.
- Kearsley, G. P. (1976). Questions and question asking in verbal discourse: A cross-disciplinary review. *Journal of Psycholinguistic Research*, 5, 355–375. doi:10.1007/BF01079934.
- LePoire, 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:10.1080/03637759909376460.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223.
- Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., et al. (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.
- 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, 28–37. doi:10.1016/j.rasd.2006.07.003.
- McDonald, S., Flanagan, S., Rollins, J., & Kinch, J. (2003). TASIT: A new clinical tool for assessing social perception after traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 18, 219–238.
- McDonald, S., Bornhofen, C., Shum, D., Long, E., Saunders, C., & Neulinger, K. (2006). Reliability and validity of the awareness of social inference test (TASIT): A clinical test of social perception. *Disability and Rehabilitation*, 28, 1529–1542.
- McLaughlin, M. L., & Cody, M. J. (1982). Awkward silences: Behavioral antecedents and consequences of the conversational lapse. *Human Communication Research*, 8, 299–316. doi:10.1111/j.1468-2958.1982.tb00669.x.
- Mesibov, G. V., Shea, V., & Schopler, E. (2007). *The TEACCH approach to autism spectrum disorders*. New York, NY: Springer.
- Paul, R. (2003). Promoting social communication in high functioning individuals with autistic spectrum disorders. *Child and Adolescent Psychiatric Clinics of North America*, 12, 87–106.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Tree, J. E. F. (2002). Interpreting pauses and ums at turn exchanges. *Discourse Processes*, 34, 37–55. doi:10.1207/S15326950DP3401_2.
- 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, 103–111. doi:10.1007/s10803-006-0160-1.

- Developmental Disorders*, 37, 1960–1968. doi: [10.1007/s10803-006-0343-3](https://doi.org/10.1007/s10803-006-0343-3).
- Webb, B. J., Miller, S. P., Pierce, T. B., Strawser, S., & Jones, W. P. (2004). Effects of social skill instruction for high-functioning adolescents with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, 19, 53–62. doi: [10.1177/10883576040190010701](https://doi.org/10.1177/10883576040190010701).
- Wechsler, D. (1999). *Manual for the Wechsler abbreviated scales of intelligence*. San Antonio, TX: Psychological Corporation.
- White, S. W., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism and Developmental Disorders*, 37, 1858–1868. doi: [10.1007/s10803-006-0320-x](https://doi.org/10.1007/s10803-006-0320-x).
- Whitehurst, G. J. (1984). Interrater agreement for journal manuscript reviews. *American Psychologist*, 39, 22–28.
- Williams, K. D., & Zadro, L. (2001). Ostracism: On being ignored, excluded, and rejected. In M. Leary (Ed.), *Interpersonal rejection*. New York, NY: Oxford University Press.
- Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, 35, 331–335. doi: [10.1007/s10803-005-3300-7](https://doi.org/10.1007/s10803-005-3300-7).