JSLHR

Research Article

Assessing Pragmatic Language in Autism Spectrum Disorder: The Yale *in vivo* Pragmatic Protocol

Elizabeth Schoen Simmons,[a] Rhea Paul,[b] and Fred Volkmar[a]

Purpose: This study compared pragmatic language in youths (9–17 years) with autism spectrum disorder (ASD) and those with typical development (TD) on the Yale *in vivo* Pragmatic Protocol (Y*i*PP), a semistructured, dynamic conversational assessment.
Method: Participants (n = 118) were divided into groups based on age and diagnosis. Each completed the Y*i*PP, which included 4 pragmatic domains (discourse management, communicative functions, conversational repair, presupposition). The participant's response to each probe was scored correct or incorrect; incorrect scores elicited cues from the examiner, and level of cue required for a correction was also scored.
Results: The Y*i*PP showed high reliability and internal consistency, with moderate concurrent validity, sensitivity, and specificity. The group with ASD performed worse overall on Y*i*PP probes compared to their TD counterparts on both error (d = 0.96) and cue (d = 0.91) scores. Item analyses revealed greater gaps between older students with ASD and their TD peers than between the 2 younger groups.
Conclusions: These data suggest that a probe measure designed to assess pragmatic abilities in children with ASD within a conversational context has some validity for contributing to diagnostic classification and can identify specific areas of pragmatic vulnerabilities as part of a clinical assessment.

A core feature, and one of the primary diagnostic symptoms, of autism spectrum disorder (ASD) is a qualitative impairment in communication (American Psychiatric Association, 2000). Current estimates suggest that a majority of individuals with this disorder function within the normal range on IQ testing (American Psychiatric Association, 2000; Dawson, Mottron, & Gernsbacher, 2008; Volkmar, Klin, & Rutter, 2005) and use spoken language as the primary means of communication. These individuals are typically referred to as having high-functioning ASD, or HFA. Research on the development of language in ASD (summarized by Kim, Paul, Tager-Flusberg, & Lord, 2014) suggests relative strengths in the areas of phonology, morphosyntax, and vocabulary when compared to pragmatic abilities.

Pragmatic language is defined by the American Speech-Language Hearing Association (ASHA) as effective and appropriate use of language to accomplish social goals, manage turns and topics in conversation, and express appropriate degrees of politeness, awareness of social roles, and recognition of others' conversational needs (ASHA, 2014). Deficits in pragmatic skills are highly prevalent in speakers with ASD (Baron-Cohen, 1988; Dewey & Everard, 1974; Kim et al., 2014; Tager-Flusberg, 1981; Volkmar, 1987), even in those functioning at the highest levels of intellectual ability (Paul, Landa, & Simmons, 2014). Pragmatic language deficits may also exist in the absence of problems with syntax, semantics, and phonology (Young, Diehl, Morris, Hyman, & Bennetto, 2005).

A variety of conversational deficits have been reported in this population (recently summarized by Paul, Landa, & Simmons, 2014), including reduced engagement in turn taking, restricted-speech acts, difficulty in making appropriate judgments about how much/little to say in conversational responses, problems in taking another's perspective in conversation, and in structuring narratives. Individuals with ASD are, in addition, prone to use of overly formal language, resulting in a pedantic style of speech that can be especially alienating to peers.

Despite the breadth of literature examining pragmatic language deficits in ASD, there is a dearth of efficient, valid, and reliable means for assessing pragmatic skills in this population (Norbury, 2014). Pragmatic skills are the most

[a]Yale University, New Haven, CT
[b]Sacred Heart University and Haskins Laboratories, Fairfield, CT
Correspondence to Elizabeth Schoen Simmons: elizabeth.schoen@yale.edu

Editor: Marilyn Nippold
Associate Editor: Martin Fujiki
Received February 10, 2014
Revision received May 22, 2014
Accepted June 12, 2014
DOI: 10.1044/2014_JSLHR-L-14-0040

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

difficult aspect of language ability to assess and quantify (Adams, Green, Gilchrist, & Cox, 2002; Landa, 2000; Young et al., 2005) and cannot be measured in the same way as other domains of language, which can be examined in socially decontextualized settings. Appropriate pragmatic function relies, and is judged, heavily on specific social-linguistic elements of the interactive context. Although several standardized instruments contain pragmatic subtests or attempt to assess this area solely, none provide a naturalistic conversational setting for doing so (Adams & Lloyd, 2005; Young et al., 2005). Instead, aspects of pragmatics are examined in single utterance items outside the context of cohesive discourse. Klin, Jones, Schultz, and Volkmar (2003) and Volkmar, Lord, Bailey, Schultz, and Klin (2004) showed that speakers with ASD tend to do better on these decontextualized, examiner-administered measures than assessment of natural conversation would predict, raising issues about the validity of the use of direct standardized assessments for measuring pragmatics in this population. And although it may be possible to identify some children with pragmatic deficits by using standardized tests (Young et al., 2005; Reichow, Salamak, Paul, Volkmar, & Klin, 2008), this identification neither provides sufficient information for planning intervention programs nor is sensitive to change during the course of intervention.

There are several questionnaires and observational measures available in the literature for assessing conversational skills (e.g., Bishop, 2006; Bishop & Adams, 1989; Brinton & Fujiki, 1993; Larson & McKinley, 1995; Paul, 2005; Prutting & Kirchner, 1983). One, the Children's Communication Checklist—Second Edition (CCC-2; Bishop, 2006), is a parent–caregiver rating scale for children ages 4–16 years that assesses several aspects of pragmatic language, including conversational discourse. Another parent report measure of language use, the Language Use Inventory (LUI; O'Neil, 2009), is aimed at younger children 18–47 months of age. It yields standardized information about the foundations of conversational development including speech acts and communicative functions. The CCC-2 and LUI both provide normative information about pragmatic language skills in children and can identify general pragmatic deficits, but, like standardized tests, are limited in their ability to provide information on specific pragmatic problems that can guide an intervention program. In addition to parent report measures of pragmatic performance, there are a handful of direct assessments available to evaluate conversational skills in spontaneous speech. The Targeted Observation of Pragmatics in Children's Conversation (TOPICC; Adams et al., 2012) is a semistructured picture-based assessment that allows the examiner to rate basic conversational skills in children. An additional direct measure, the Peanut Butter Protocol (Creaghead, 1984), was designed for very young children, ages 3–5 years, to observe communicative intents and conversational devices while sharing a snack. While these direct measures both offer insight into a child's conversational repertoire, they do not offer normative information and are only useful at restricted age ranges.

Thus, while there are some instruments that supply standardized information on pragmatic performance, the majority of currently available direct measures of pragmatic functioning provide neither psychometric information nor a quantitative metric of pragmatic competence in specific domains that can be used to establish a level of baseline function or document change in intervention (Brinton, Robinson, & Fujiki, 2004; Landa, 2000; Norbury, 2014).

One solution to this problem is to make use of dynamic assessment procedures. Dynamic methods are contrasted with *static* forms of assessment, such as standardized tests, which describe current level of performance by holding contextual support to a minimum. Dynamic assessment is designed to systematically manipulate the degree of environmental support for task completion so that a child's optimal level of achievement can be observed and the amount of contextual structure needed to enable that optimal level can be identified (Gutiérrez-Clellen & Peña, 2001; Swanson & Lussier, 2001). Dynamic assessment can incorporate instruction and feedback into the testing process, targeted to an individual's performance (Lidz & Peña, 1996; Paul & Norbury, 2012). It is especially valuable for evaluating growth over time, or in a treatment program, since it allows the determination of whether, with time or treatment, lower levels of environmental support are needed to produce a target behavior than they were at the outset. Dynamic assessment appears ideally suited to the evaluation of conversational competence since it allows for observations within a natural discourse context, obviates the problem of a lack of normative reference data for conversational skills by providing a measure of change over time, and enables the observation of a wide range of conversational behaviors.

In the present report, we employed both dynamic and static measures of conversational ability to address the following research questions:

1. Do scores on our naturalistic assessment measure differ significantly between children and adolescents with ASD and their counterparts with typical development (TD), and in what areas of pragmatics?
2. Can the instrument provide significant discrimination between the diagnostic groups?
3. Does the instrument meet psychometric standards of reliability, internal consistency, validity, sensitivity, and specificity?

We also consider whether the use of the dynamic scoring system built into this instrument adds to the ability to use it for intervention planning and as a measure of response to intervention.

Method

Participants

Participants were recruited from prior participation in research at the university's laboratory, through community resources, and via electronic media. The Yale University

School of Medicine's Institutional Review Board approved the study protocol.

School-age children and adolescents, ages 9–17 years, participated in this research. Group assignments were based on clinical diagnosis and participant age, as follows:

- The HFA-Younger (HFA-Y) group: ages 9–12 years with a diagnosis of high-functioning autism spectrum disorder (n = 47).
- The HFA-Older (HFA-O) group: ages 13–17 years with the same diagnosis (n = 30).
- The TD-Younger (TD-Y) group: ages 9–12 years with no family history of ASD and/or psychiatric, learning, or neurological disorder present in the child (n = 26).
- The TD-Older (TD-O) group: ages 13–17 years with no family history of ASD and/or psychiatric, learning, or neurological disorder present in the child (n = 15). This was the smallest of the four groups. Difficulties recruiting these participants were primarily due to our lengthy research protocol requiring two full days, which older students with TD found difficult to accommodate in their schedules. Power analyses indicate power of greater than 80% to detect differences in overall error and overall cue scores using the present sample size.

Inclusionary/exclusionary criteria. All participants had full-scale IQs of at least 70 (Gillberg & Ehlers, 1998; Siegel, Minshew, & Goldstein, 1996) and fluent expressive language, as validated by a standard Core Language score ≥70 on the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF–4; Semel, Wiig, & Secord, 2003). Participants with typical development were matched to those with HFA on the basis of chronological age and full-scale IQ. There were no significant differences between mean full-scale IQ scores of HFA and TD participants. Male to female ratios in the sample with ASD match those reported in the literature (Fombonne, 2003). While gender was closely matched across diagnostic groups, there was a higher percentage of females in the TD-O group compared to the TD-Y group (see Table 1 for gender distribution).

Participants in the TD group were screened for other psychiatric disorders using the Child and Adolescent Symptom Inventory—Fourth Edition (Gadow & Sprafkin, 2002). The Social Communication Questionnaire—Current and Lifetime Forms (SCQ; Rutter, Bailey, & Lord, 2003) were completed by a caregiver to ensure absence of autism symptomatology in TD participants. Exclusion criteria for both groups included the presence of any known seizure activity or medication that could have an impact on speech production. Individuals with cerebral palsy, tuberous sclerosis, hemiparesis, ataxia, Fragile X or other chromosome abnormality, orofacial abnormalities (e.g., cleft palate), or gross-motor or uncorrected sensory impairments (e.g., hearing loss) were excluded from participation.

Confirmatory diagnosis of ASD. The use of the broad ASD diagnostic category follows the diagnostic scheme utilized in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM–IV–TR*; American Psychiatric Association, 2000) and includes high-functioning autism, Asperger syndrome, and pervasive developmental disorders–not otherwise specified (PDD-NOS). All participants in the HFA groups had been previously diagnosed with an ASD. Confirmatory diagnosis was completed at the time of study participation. Each participant in the HFA groups met the *DSM-IV-TR* (American Psychiatric Association, 2000) criteria for one of the autism spectrum disorders as judged by at least two experienced clinicians. Diagnoses were confirmed by scores on both the Autism Diagnostic Interview—Revised (ADI–R; Rutter, LeCouteur, & Lord, 2003) and the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2001) revised diagnostic algorithms as part of study participation. Both instruments were administered by trained research clinicians.

Procedure

Standardized measures. The CELF–4 was used to measure expressive and receptive language skills in the areas of syntax, morphology, and vocabulary. All participants completed standardized measures of IQ and language functioning. The HFA groups were administered the Differential Abilities Scale—II: School-Age Version (DAS–II: SA; Elliot, 2006) as a measure of full-scale IQ. Given there were no concerns relative to cognition for the TD groups, the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) was administered as a brief measure of cognition to assess IQ for these participants. Table 1 provides characterization information on participants. No significant differences in performance were observed between the HFA-Y versus TD-Y groups or between HFA-O and TD-O groups on either the CELF Core Language Standard Scores or measures of full-scale IQ.

Yale in vivo *Pragmatic Protocol.* Each participant completed the Yale *in vivo* Pragmatic Protocol (Y*i*PP). The Y*i*PP is a structured conversational task designed to appear naturalistic to the participant. Portions of the activities used as part of this protocol were adapted from Creaghead's (1984) Peanut Butter Protocol. Specific items on the Y*i*PP were extensively piloted using an iterative design process in which specific pragmatic probes were adapted based on participant performance and examiner feedback (Chuba, Paul, Klin, & Volkmar, 2003; Loomis, Schoen, & Paul, 2011; Miles, Paul, Klin, Volkmar, & Cohen, 2001; Schoen, & Paul, 2009). Theoretical bases for Y*i*PP domains were drawn from our earlier work on pragmatic classification schemes (Paul, 2005; Paul & Fahim, 2014; Weiss, Paul, & Norbury, 2006), which are, in turn, derived from Chapman (1981). Domains of the Y*i*PP were aligned with the ASHA definition cited earlier by operationalizing each aspect of the definition as a domain of Y*i*PP items:

- Language to accomplish social goals: Examines ability to express a range of communication functions
- Manage turns and topics in conversation: Examines discourse management in terms of topic initiation, termination, maintenance, and turn taking

Table 1. Characteristics of participant groups.

	HFA (n = 77)		TD (n = 41)	
	HFA-Y (n = 47)	HFA-O (n = 30)	TD-Y (n = 26)	TD-O (n = 15)
Variable	***M (SD)***	***M (SD)***	***M (SD)***	***M (SD)***
CA in years	10.69 (1.19)	15.22 (1.49)	11.16 (1.19)	14.75 (0.80)
Standardized testing measures[a]				
Full-scale IQ	99.50 (19.55)	99.50 (21.68)	106.81 (10.24)	102.60 (10.32)
CELF–4 Core Language	95.74 (22.15)	95.50 (22.95)	107.50 (9.23)	105.53 (10.77)
Gender				
Male	41 (87%)	26 (87%)	21 (81%)	11 (73%)
Female	6 (13%)	4 (13%)	5 (19%)	4 (27%)
Maternal education (≥ 4 year college)	68%	63%	65%	80%

Note. HFA (Y or O) = high-functioning autism (younger or older); TD (Y or O) = typically developing (younger or older); CA = chronological age; CELF–4 = Clinical Evaluation of Language Fundamentals—Fourth Edition (Semel, Wiig, & Secord, 2003).

[a]HFA groups completed the Differential Abilities Scale—II: School-Age (DAS–II:SA; Elliot, 2007); TD groups completed Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). All assessments had a standard score of 100 and standard deviation of 15.

- Express appropriate degrees of politeness and awareness of social roles: the original Y*i*PP examined register variation in speech to varying interlocutors and use of politeness variations, but these items were eliminated as not being discriminative and because the participants found them unnatural, through the iterative development process during the pilot phase
- Recognition of others' conversational needs: recognition of interlocutor's informational needs is examined through presuppositional items; need for clarification through conversational repair items

The Y*i*PP contains a series of probes to collect information on a variety of conversational speech acts. Within this 30-min conversation, the examiner follows a script that inserts 19 pragmatic probes to elicit a target behavior within four conversational domains (see Table 2). The protocol was designed to be naturalistic and child directed (e.g., participant chooses a topic to discuss) and attempted to include tasks and activities that school-age participants encounter on a daily basis.

*Yi*PP *administration.* The participant and examiner were seated across from each other, and the participant was told that he or she was going to have a brief conversation with the examiner. If the participant produced a targeted response to the probe, the examiner continued the conversation. If the participant did not respond to the probe, the examiner provided a series of prompts to determine whether, and at what level of explicitness, scaffolding was helpful in eliciting the target pragmatic behavior. The cues were administered in a hierarchical fashion with the examiner providing the lowest level of support (e.g., the cue with the least amount of information) to the participant first. More supportive cues were provided if the participant failed to respond to the previous ones. Cues were provided in this sequence until the participant produced the target pragmatic behavior or until the final cue was administered. The cues were presented in the following order: expectant waiting, gesture/facial expression, nonspecific verbal cue (e.g., "What?") or repetition of pragmatic probe, and specific verbal cue. The examiner script and list of specific verbal cues are included in the online supplemental materials. Each Y*i*PP was video recorded for coding at a later date. At the conclusion of the Y*i*PP, the participant was debriefed if necessary.

Coding. Coding was completed by a trained research assistant, who was blind to participant diagnosis, from video recorded protocols. A total of 19 pragmatic language acts in response to the 19 pragmatic probes administered during the Y*i*PP were coded. Each pragmatic language act produced in response to a probe was assigned an error score and a cue score. The error score recorded whether a response was correct/incorrect. Error scores ranged from 0 through 2, with a score of 0 representing the best possible performance (e.g., a correct, appropriate pragmatic response), a score of 1 representing a mildly inappropriate response, and a score of

Table 2. Pragmatic behaviors probed by Yale *in vivo* Pragmatic Protocol (Y*i*PP).

Conversational domain	Number of pragmatic probes	Target pragmatic behavior (corresponding script number)
Discourse management	6	Initiation (1), request information (2), provide background information (5), terminate topic (13), topic maintenance (15), turn taking (14)
Communicative functions	4	Hypothesizing (3), commenting (16), requesting (17), protesting/denial (18)
Conversational repair	4	Requesting clarification (4, 7, 8, 12)
Presupposition	5	Comment contingently (6), ambiguous article (9), too little information (10, 19), ambiguous pronoun (11)

2 representing a clearly inappropriate or no response. The cue score measured the level of cueing necessary for the participant to produce the target pragmatic language act, if the initial response was incorrect or absent. Cue scores ranged from 0 through 6, with 6 indexing an appropriate, spontaneous response with no cue; a score of 0 indicated no response regardless of level of cueing provided. Thus, lower error scores were indicative of better performance on the Y*i*PP, whereas lower cue scores were indicative of *worse* performance. Table 3 outlines the coding scheme used for scoring pragmatic language acts occurring during the Y*i*PP. If the examiner failed to administer a given pragmatic probe or cue that item was discarded from coding and analysis. Approximately 2% of the total number of probes presented were spoiled and eliminated from analysis. A full description of coding procedures is provided in the online supplemental materials.

Data Analysis

To answer the first research question, descriptive statistics were computed for error and cue scores for each of the four pragmatic domains measured by the Y*i*PP for each of the subgroups (HFA-Y, HFA-O, TD-Y, TD-O) by summing the error scores for each probe, dividing the sum by the number of probes in the given domain for each participant, then averaging across participants. Mean cue scores for each domain were calculated the same way. Means and standard deviations were also calculated for a total error score by summing the mean error scores from each of the four pragmatic domains then dividing the sum by four. The total cue score was calculated in the same fashion. One-way analysis of variance (ANOVA) and planned post hoc comparisons were used to assess group differences on error and cue scores for each domain and for total error and total cue scores. Pairwise comparisons between diagnostic groups for the younger (HFA-Y vs. TD-Y) and older (HFA-O vs. TD-O) participants were conducted using Tukey's honestly significant difference (HSD) to correct for multiple comparisons in the case where variables were homogenous. For variables with a statistically significant Levene statistic (e.g., $p < .05$) indicating lack of homogeneity of variance, Games–Howell pairwise comparison was employed. Effect sizes for significant differences are also reported.

For the second research question, discriminant function analysis was performed to validate the identification of pragmatic language impairment in individuals with HFA in order to evaluate the diagnostic utility of the Y*i*PP. This methodology of group classification was chosen over linear regression given the homogeneity of variance of predictor variables and our intention to use the Y*i*PP to make a dichotomous judgment between the presence and absence of pragmatic impairment.

To answer the third research question, several basic psychometric properties of Y*i*PP scores were examined. Our last question was addressed by means of item analysis comparing error versus cue scores.

Results

Interrater Reliability

Interrater reliability was established using the kappa statistic to measure agreement between the primary Y*i*PP coder and a second trained coder (e.g., first author of this article). A randomly selected 10% of Y*i*PP videos were recoded by the second trained coder. For codes where there was disagreement between the two coders, those video segments were re-watched, and a consensus code was achieved and used for the final data set. The kappa statistic between raters was .81 (very high; Landis & Koch, 1977) and .71 (substantial; Landis & Koch, 1977) for error scores and cue scores, respectively.

Internal Consistency

Mean error scores from discourse management, communicative functions, conversational repair, and presupposition domains, in addition to the total error score, all demonstrated high consistency, Cronbach's $\alpha = .82$ (excellent; George & Mallery, 2003). Corrected item total correlations were computed for each error score at the item level and revealed removal of any given item would decrease the alpha value. Similar consistency was observed between the mean cue scores for each of the domains including the total

Table 3. Y*i*PP coding rubric.

Error score	Description	Example of responses during request clarification probe: Muffled speech
2	Incorrect/no response	You shouldn't cough.
1	Mildly inappropriate, unusual response	What?
0	Correct/appropriate response	Can you repeat yourself?
Cue score		
6	Appropriate, spontaneous response; error score of 0; no cue given	
5	Mildly inappropriate response; error score 1; no cue given	
4	Expectant waiting	
3	Gesture/facial expression	
2	Nonspecific verbal cue/repetition	
1	Specific verbal cue	
0	No response to any prompt	

cue score, Cronbach's α = .83. These data suggest the four domains represent, to some degree, the same construct and indicate acceptable internal consistency for a set of subscales within one measure. Domain level Cronbach's alphas are reported in Table 4.

Concurrent Validity

Concurrent validity of the Y*i*PP was assessed by correlating the total error and total cue scores from each participant with the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005) total score derived from parent- and teacher-completed forms. The SRS is a 65-item Likert scale questionnaire designed to be completed by caregivers and educators. It provides information on social functioning, including pragmatic language (e.g., conversation) and autism symptomotology (e.g., stereotyped and repetitive movements). The SRS was chosen as a measure of concurrent validity because it has been demonstrated to be a psychometrically strong measure of social function (Constantino et al., 2003); moreover, it has good concordance with the Y*i*PP since a subset of questions from the SRS yield information about pragmatic language. The SRS yields a T-score with a higher score indicating greater autism symptomotology. The total error score from the Y*i*PP was positively correlated with both the SRS total score from parent and teacher reports with Pearson's $r = .41$ and .31, $p < .001$, respectively, indicating that the poorer the pragmatic performance, as measured by higher error scores, the more autism-like features demonstrated by the participant as measured by higher SRS total score. The total cue score was negatively correlated with both the SRS total score from parent and teacher reports with Pearson's $r = –.39$ and –.36, $p < .001$, respectively. This suggests that the more cueing required, the higher the SRS total score, suggesting greater autism symptoms.

Group Differences Within Four Pragmatic Categories

One-way ANOVAs were computed to examine the differences among the four groups on error and cue scores in the four broad pragmatic domains (discourse management, communicative functions, conversational repair, presupposition) and total error and total cue scores. ANOVA results revealed significant differences for both error and cue scores between HFA and TD groups on all but the communicative functions domain. Significant differences between HFA and TD groups were also identified on the total error score and total cue score. ANOVA results are outlined in Table 5.

Table 4. Cronbach's alphas for domain-level error and cue scores.

Pragmatic domain	Cronbach's α	
	Error score	Cue score
Discourse management	0.86	0.82
Communicative functions	0.71	0.74
Conversational repair	0.83	0.85
Presupposition	0.85	0.85

Item Analyses

Pairwise comparisons examining group differences in responses to each Y*i*PP item within the four pragmatic domains, along with mean (*SD*) scores for each group on each item, appear in Tables 6 (error scores) and 7 (cue scores).

Discourse management. Both HFA groups (HFA-Y and HFA-O) performed more poorly on this domain compared to their age-matched typically developing peers, as evidenced by their higher mean error score and lower mean cue score (see Tables 6 and 7), with large effect sizes. The HFA-Y group produced more pragmatic errors on the background information and topic maintenance probes compared to the TD-Y group. The HFA-Y group also required significantly more cueing to produce a pragmatic response on the request information probe. The HFA-O group demonstrated more pragmatic errors compared to age-matched peers on a single probe within this domain: the request information probe. Similar to their younger counterparts, the HFA-O group required greater cueing to produce a pragmatic response on the request information and topic maintenance probes.

Conversational repair. The significant differences found in the overall ANOVA comparing error scores for HFA and TD groups on this domain appears, from the pairwise comparisons in Table 6, to be attributable entirely to the difference in just one item: decreased volume. The younger HFA group produced more pragmatic errors in response to the decreased volume probe compared to the age-matched TD group, and also required greater cueing to produce a response on this probe. The HFA-O group showed the same pattern of response to decreased volume. In addition, they required significantly more cueing on the ambiguous statement probe.

Presupposition. Significant differences were observed between the HFA and TD groups on both the mean error score and the mean cue score for this domain, with large effect sizes for both age subgroups. Significantly different error scores were seen on the ambiguous pronoun item in both age groups, as well as on the comment contingently and too little verbal information items in the older groups only. The HFA-O group also required more clinician cuing on both these probes.

Communicative functions. Post hoc comparisons were not conducted on this domain, as a one-way ANOVA revealed no significant differences between HFA and TD groups on mean error and cue scores for this domain.

Clinical Properties

An additional aim of this study was to validate the identification of pragmatic language impairment in individuals with autism whose basic language performance falls within the average range on standardized testing. These individuals may fail to qualify for speech-language services

Table 5. Analyses of variance (ANOVAs) between HFA and TD groups for pragmatic domains.

Pragmatic domain	Mean score (*SD*) HFA group (*n* = 77)	Mean score (*SD*) TD group (*n* = 41)	Significance (*p*)*	*d* (effect size)[a]
Error scores				
Discourse management	0.71 (0.37)	0.34 (0.28)	0.001	1.15
Communicative functions	0.72 (0.42)	0.73 (0.48)	*ns*	*ns*
Conversational repair	0.82 (0.50)	0.51 (0.35)	0.001	0.72
Presupposition	0.83 (0.37)	0.48 (0.34)	0.001	0.97
Total error score	0.77 (0.30)	0.52 (0.24)	0.001	0.96
Cue scores				
Discourse management	4.39 (0.98)	5.25 (0.68)	0.001	0.88
Communicative functions	4.08 (1.32)	4.19 (1.07)	*ns*	*ns*
Conversational repair	3.76 (1.45)	4.70 (1.13)	0.001	0.72
Presupposition	3.42 (1.18)	4.44 (1.29)	0.001	0.83
Total cue score	3.91 (0.90)	4.65 (0.72)	0.001	0.91

[a]>.70 = medium effect size, >.80 = large effect size (Cohen, 1988).

*$p < .05$, one-tailed.

without a quantitative assessment of the relevant parameters. Since the participants with ASD achieved significantly worse total error and total cue scores compared to their typically developing counterparts, discriminate function analyses were conducted to learn if these scores could predict group membership.

Discriminant function analyses were performed using total error score, total cue score, or total error score + total cue score for each age group to assess the utility of these scores as a predictor of membership in the diagnostic group clinically assigned (e.g., HFA or TD). These scores were chosen as they provided a global rating of pragmatic functioning across the four broad domains assessed by this instrument, and these scores differed significantly between diagnostic groups as revealed by ANOVA results in Table 5. The total error score correctly classified 60.3% of the 9- to 12-year-old cohort (canonical correlation = .33, $p < .005$) and 78% of the 13- to 17-year-old cohort (canonical correlation = .52, $p < .001$) into their diagnostic group. Discriminant function analyses of the total cue score accurately classified

Table 6. Summary of error scores for pragmatic domains and individual probes.

Pragmatic domain/probe	Younger groups (*n* = 73) Mean error score (*SD*) HFA-Y (*n* = 47)	Younger groups (*n* = 73) Mean error score (*SD*) TD-Y (*n* = 26)	Younger groups (*n* = 73) Significance (*d* effect size)[b]	Older groups (*n* = 45) Mean error score (*SD*) HFA-O (*n* = 30)	Older groups (*n* = 45) Mean error score (*SD*) TD-O (*n* = 15)	Older groups (*n* = 45) Significance (*d* effect size)
Discourse management	0.72 (0.37)	0.34 (0.30)	0.001 (1.13)	0.71 (0.38)	0.36 (0.25)	0.005 (1.09)
Initiation	0.96 (0.59)	0.72 (0.52)	*ns*	0.97 (0.31)	0.39 (0.10)	*ns*
Request information	0.68 (0.74)	0.29 (0.46)	*ns*	0.62 (0.57)	0.08 (0.28)	0.002 (1.20)
Background information	1.02 (0.68)	0.35 (0.56)	0.001 (1.08)	0.85 (0.67)	0.53 (0.51)	*ns*
Termination	0.64 (0.72)	0.38 (0.57)	*ns*	0.46 (0.51)	0.15 (0.38)	*ns*
Response to cues to change speakers	0.35 (0.64)	0.15 (0.37)	*ns*	0.28 (0.46)	0.12 (0.33)	*ns*
Topic maintenance	0.66 (0.73)	0.23 (0.43)	0.01 (0.72)	0.63 (0.61)	0.24 (0.56)	*ns*
Communicative functions	0.68 (0.42)	0.72 (0.53)	[a]	0.79 (0.39)	0.73 (0.40)	[a]
Conversational repair	0.84 (0.53)	0.58 (0.36)	*ns*	0.79 (0.46)	0.40 (0.32)	0.03 (0.98)
Muffled speech	0.85 (0.78)	0.64 (0.64)	*ns*	0.85 (0.67)	0.53 (0.62)	*ns*
Decreased volume	0.76 (0.71)	0.35 (0.56)	0.04 (0.64)	0.58 (0.66)	0.12 (0.33)	0.01 (0.88)
Unfamiliar acronym	0.60 (0.74)	0.46 (0.65)	*ns*	0.67 (0.74)	0.41 (0.51)	*ns*
Ambiguous statement	1.16 (0.71)	0.91 (0.73)	*ns*	1.03 (0.70)	0.56 (0.81)	*ns*
Presupposition	0.82 (0.40)	0.55 (0.34)	0.01 (0.73)	0.85 (0.35)	0.38 (0.33)	0.001 (1.38)
Comment contingently	0.09 (0.35)	0.12 (0.33)	*ns*	0.24 (0.50)	0.00 (0.00)	0.43 (0.68)
Ambiguous article	1.22 (0.67)	0.94 (0.42)	*ns*	1.00 (0.58)	0.55 (0.69)	*ns*
Too little verbal information	1.07 (0.74)	0.84 (0.85)	*ns*	1.22 (0.66)	0.41 (0.71)	0.002 (1.18)
Ambiguous pronoun	1.18 (0.78)	0.67 (0.57)	0.14 (0.75)	1.31 (0.59)	0.65 (0.86)	0.041 (0.89)
Too little written information	0.73 (0.85)	0.31 (0.62)	*ns*	0.39 (0.66)	0.29 (0.59)	*ns*
Total mean error score	0.77 (0.31)	0.55 (0.28)	0.011 (0.74)	0.78 (0.28)	0.47 (0.67)	0.002 (0.60)

[a]Post hoc testing not completed due to a nonsignificant ANOVA. [b]>.60 = small effect size, >.70 = medium effect size, >.80 = large effect size (Cohen, 1988).

Table 7. Summary of cue scores for pragmatic domains and individual probes.

Pragmatic domain/probe	Younger groups (n = 73)			Older groups (n = 45)		
	Mean cue score (SD)		Significance (d effect size)[b]	Mean cue score (SD)		Significance (d effect size)
	HFA-Y (n = 47)	TD-Y (n = 26)		HFA-O (n = 30)	TD-O (n = 15)	
Discourse management	4.38 (1.06)	5.19 (0.72)	0.002 (0.89)	4.42 (0.88)	5.34 (0.62)	0.004 (1.24)
Initiation	3.49 (2.12)	3.48 (2.16)	*ns*	2.78 (1.84)	4.12 (1.73)	*ns*
Request information	4.24 (2.11)	5.38 (1.02)	0.03 (0.69)	4.62 (1.66)	5.85 (0.56)	0.008 (0.99)
Background information	3.91 (1.98)	5.00 (1.83)	*ns*	4.25 (2.03)	5.06 (1.44)	*ns*
Termination	4.70 (1.72)	5.42 (1.24)	*ns*	5.18 (1.19)	5.77 (0.60)	*ns*
Response to cues to change speakers	5.39 (1.39)	5.69 (0.88)	*ns*	5.47 (1.02)	5.88 (0.33)	*ns*
Topic maintenance	4.77 (1.72)	5.46 (1.27)	*ns*	4.41 (1.88)	5.59 (1.06)	0.035 (0.77)
Communicative functions	4.13 (1.44)	4.17 (1.10)	[a]	4.01 (1.16)	4.22 (1.05)	[a]
Conversational repair	3.66 (1.40)	4.39 (1.25)	*ns*	3.90 (1.52)	5.18 (0.76)	0.009 (1.07)
Muffled speech	3.94 (2.16)	3.96 (2.15)	*ns*	4.18 (1.85)	4.88 (1.58)	*ns*
Decreased Volume	3.87 (2.07)	5.12 (1.61)	0.031 (0.67)	4.24 (2.17)	5.76 (0.75)	0.004 (0.94)
Unfamiliar acronym	4.64 (2.05)	4.85 (1.95)	*ns*	4.45 (2.18)	5.35 (1.22)	*ns*
Ambiguous statement	2.19 (2.17)	3.64 (2.22)	*ns*	2.81 (2.07)	4.67 (2.23)	0.035 (0.86)
Presupposition	3.48 (1.20)	4.12 (1.31)	*ns*	3.33 (1.15)	4.93 (1.10)	0.001 (1.42)
Comment contingently	5.72 (1.08)	5.50 (1.42)	*ns*	5.21 (1.76)	6.00 (0.00)	*ns*
Ambiguous Article	2.14 (1.99)	2.72 (1.93)	*ns*	3.00 (2.06)	4.64 (1.86)	*ns*
Too little verbal information	2.89 (2.27)	3.52 (2.52)	*ns*	2.13 (2.11)	4.71 (2.20)	0.001 (1.19)
Ambiguous pronoun	2.07 (2.19)	3.13 (2.44)	*ns*	1.75 (1.78)	4.00 (2.71)	0.030 (0.98)
Too little written information	4.05 (2.48)	4.92 (2.17)	*ns*	4.67 (2.19)	5.12 (1.83)	*ns*
Total mean cue score	3.91 (0.91)	4.47 (0.80)	0.037 (0.65)	3.91 (0.90)	4.92 (0.48)	0.001 (1.41)

[a]Post hoc testing not completed due to a nonsignificant ANOVA. [b]>.70 = medium effect size, >.80 = large effect size (Cohen, 1988).

62% of younger participants (canonical correlation = .30, $p < .01$) and 74% of older participants (canonical correlation = .53, $p < .001$) into their correct diagnostic group.

Using both total error score and total cue score in the discriminant function model did not improve correct classification for the younger participants (60.3% correctly classified; canonical correlation = .33, $p < .02$) compared to the total error score alone. For the older group of participants, using both scores (76% correctly classified; canonical correlation = .54, $p < .001$) did not improve the accuracy of classification using either total error score alone and was less accurate than using total cue score alone.

Sensitivity and specificity were computed for total error score and total cue score for both older and younger age groups. These are reported in Table 8; sensitivity and specificity were derived from data of our typically developing participants based on means and standard deviations. The cut-off score was calculated by adding one standard deviation to the mean total error score for both the younger and older typically developing groups. If a participant received a score above this cut-off point, the participant was classified as HFA. If the participant received a score below this cut-off point, the participant was classified as TD. The same procedure was used for the total cue score. The decision to use one standard deviation to determine the cut-off score was reached upon evaluation of the criteria for research determination for inclusion in other studies of various types of language disorder (e.g., Redman & Rice, 2002; Rice & Buhr, 1992; Rice, Wexler, & Hershberger, 1998).

These cut-off points were validated using the receiver operating characteristic (ROC) curve. The area under the curve, a measure of diagnostic accuracy, was also obtained. An area under the curve less than .70 is considered poor diagnostic test accuracy; areas between .70 and .80 are deemed

Table 8. Sensitivity and specificity of summary scores.

Group	Cut-off score	Sensitivity	Specificity	Area under ROC curve	LR+	LR–	PPV	NPV
Younger participants								
Total error score	0.83	36.17%	84.62%	0.88	2.25	0.88	0.81	0.42
Total cue score	5.31	68.09%	50.00%	0.67	1.36	0.64	0.68	0.71
Older participants								
Total error score	0.64	66.67%	88.24%	0.86	5.58	0.38	0.92	0.58
Total cue score	5.38	96.97%	17.64%	0.85	1.18	0.16	0.71	0.75

Note. ROC = receiver operating characteristic; LR+ = positive likelihood ratio; LR– = negative likelihood ratio; PPV = positive predictive value; NPV = negative predicative value.

fair diagnostic test accuracy, and .81 and .90 are considered good diagnostic test accuracy (Feldman et al., 2005). With the exception of the total cue score for the younger participants, area under the curve for the remaining variables had a range of .85–.88, suggesting the Y*i*PP has good diagnostic accuracy (see Table 8). Similarly, the total error score provides a measure of specificity, while the total cue score provides a measure of sensitivity. When used together, both the total error and total cue score appear to provide useful information in the identification of pragmatic impairments. Positive and negative likelihood ratios, as well as predictive values, are also reported in Table 8. The positive predictive value (PPV) provided a measure of true positives while the negative predictive value (NPV) provided a measure of true negatives.

Discussion

These data suggest that both preadolescents and adolescents with ASD show significantly more errors on probe items involving areas of discourse management, conversational repair, and presupposition than peers with TD. When errors are made, those with ASD generally require higher levels of cueing in these areas to correct their errors than do peers. Several items designed to probe discourse management account for these higher levels of error. In the preadolescent group, providing background information and topic maintenance account for the difference; in the older group, only requesting information does. For presupposition, errors in understanding ambiguous pronouns are more frequent at both ages; adolescents also comment contingently less often than peers. In the area of conversational repair, only decreased volume differentiates the diagnostic groups at both age levels.

Item Analysis

One can examine item-level performance across the two types of scores for the two groups at the two age levels in an effort to understand the differences found. This examination reveals that for discourse management, failure to find significant differences between preadolescents with and without ASD on requesting information is due to the fact that the younger TD children are making more errors than their older counterparts and are therefore performing more similarly to the group with ASD. By adolescence, typical youths make virtually no errors on this item, while participants with ASD continue to make the same type of error as younger counterparts. In terms of topic maintenance, the significant difference between groups at the younger age is likely due to the larger sample size of this comparison, since the actual scores of the two older groups are virtually unchanged from their respective scores at preadolescence. Supporting this notion is the fact that there is a significant difference in error scores between the older groups, suggesting that the few times errors are made by youths with TD, they need only the lowest level of support (mean score of 5.6) to correct themselves, whereas the younger ASD participants need slightly more cueing, resulting in a gap between older speakers with ASD and TD that is larger than in the younger groups on this measure. For background information, there does seem to be some closing of the gap between diagnostic groups from the younger to the older age level, suggesting participants with ASD may be improving in this skill with age.

In examining conversational repair scores, it can be seen that the only item, decreased volume, differentiates the groups. Examining the scores suggests that speakers with ASD show a trend toward making more errors on all items than TD peers, but for decreased volume the TD youths are making fewer errors than on any of the other items. It would appear that this item is as hard as the others for those with ASD but easier for the TDs. It may be that this one item would be sufficient as a marker of need for assistance with conversational repair in future iterations of the Y*i*PP protocol.

On the presupposition items, performance on commenting contingently shows large variability for both groups at the preadolescent level, which may account for the failure of this difference to reach significance. Adolescents with TD make no errors at all on this item, so they show no variability, while the standard deviation in the ASD group remains relatively large. Thus, on this item, the performance of the TD adolescents appears to have both improved and stabilized, while individuals with ASD continue to show variable performance, resulting in a significant difference at the older level, while variability in both groups limits the power to find a difference at the younger age level. Performance on the ambiguous pronoun item shows more errors in the ASD groups at both age levels, but error score differences are significant only for older participants. Again, this appears to be an issue of an increasing gap in performance due to a need for much less cueing by adolescents with TD (on average, Level 4 cues) than younger counterparts (on average, Level 3). Participants with ASD continue to need extensive cueing to correct errors at both age levels.

Pragmatics and prosody. This examination suggests that, generally, youths with ASD make more errors and need more extensive cues to correct errors than peers with TD at both age levels. There is a trend for the adolescents with TD to show both decreases in errors and increases in response to cues on this protocol, which is not seen in the group with ASD, although these trends are somewhat item specific. Interestingly, this trend parallels our recent report on the development of prosody in this population (Lyons, Simmons, & Paul, 2014), in which we found the same pattern (i.e., significant differences between ASD and TD participants were seen more broadly in adolescent than in preadolescents, as the typically developing group advanced in their mastery of prosody, and the group with ASD remained relatively stagnant in both these areas). The link between pragmatic and prosodic development in this study is most evident in the finding that the ASD group was minimally able to repair most conversational breakdowns based on verbal content, but had significantly more difficulty repairing breakdowns based on paralinguistic signals, a

finding that replicates one in an earlier study using a natural conversational sample (Paul, Orlovski, Marcinko, & Volkmar, 2009). Again, the convergence of these findings emphasizes the importance of addressing both linguistic and paralinguistic aspects of social communication in developing intervention.

Diagnostic utility. Although the discriminant function analysis, studies of sensitivity and specificity, likelihood ratios, and predictive values for the Y*i*PP show only fair levels of accuracy, they do suggest that most students with ASD in the 9–17 year age range will be accurately classified by this measure. Moreover, it is unlikely that the Y*i*PP will be used primarily to diagnose ASD. Rather, we envision that it may be useful, not only to identify pragmatic deficits in ASD, but to serve as an index of pragmatic skill in students whose diagnoses may be more borderline or difficult to determine. For example, the *DSM–5* (American Psychiatric Association, 2013), in addition to modifying diagnostic classifications for ASD, has also added a new diagnostic category, social (pragmatic) communication disorder, characterized by the presence of pragmatic difficulties in the absence of autistic (primarily repetitive behavior) symptoms. Although there is a good deal of controversy in the field regarding the characterization and differentiability of this disorder (see Paul & Norbury, 2012; Norbury, 2014), its presence in the *DSM–5* will doubtlessly lead to the necessity of a valid measurement of pragmatic function, not only in children with ASD, but in those with other social and communication disorders. Similarly, it has long been observed that children with nonautistic language disorders, as well as those with learning disabilities, exhibit problems with pragmatic uses of language, even though they may differ in quality and severity from those seen in ASD (e.g., Bryan, Burstein, & Ergul, 2004; Paul & Norbury, 2012; Weiner, 2002). Having an additional tool to document these in natural contexts could contribute to clinicians' ability to assess and plan treatment for these aspects of the students' disability.

Clinical implications. The present findings suggest the widening of an "achievement gap" in these critical social communication skills during the adolescent years, when typically developing youths show a dramatic transition to more peer-focused, complex social interactions that require a range of new skills for their negotiation. This observation suggests that older children and adolescents with HFA, rather than growing out of their need for support within the social aspects of language, may require additional services during this period in order to help narrow the gulf between them and peers with whom they may aspire to interact. A tool like the Y*i*PP may be helpful in structuring these services and supports.

Bellini, Gardner, and Markoff (2014) have argued that one reason social skills training has shown such low levels of efficacy in meta-analytic studies is that it is often delivered without any assessment of the client so that specific areas of strength and weakness are not identified and used to guide the development of an intervention program. Although the study of efficacy in pragmatic language intervention is in its infancy, as evidence-based intervention programs are developed and tested, having an assessment tool that can guide their individualization may help to improve their efficacy. Similarly, clinicians working with clients with HFA may be able to use data they collect from probe measures like the Y*i*PP to identify specific pragmatic deficits and develop individually targeted pragmatic intervention programs.

Study limitations. Modest sample size may have limited our ability to find stronger diagnostic support for the Y*i*PP. It is noteworthy, however, that significant differences were found more frequently between the two older groups, despite the limited size of the older TD sample. Gathering data on larger samples, including narrower age brackets, would certainly add to the diagnostic utility of the measure, and determine whether or not a probe like this should be limited in usage to the adolescent age group, when the gap in performance appears widest. Similarly, refinement of item inclusion, focusing on items that discriminate at particular age levels, eliminating nondiscriminating items, and adding items to increase discriminability for domains (e.g., communicative functions) or subdomains (e.g., verbal vs. paralinguistic forms of communication repair) would increase the Y*i*PP's power to serve as a basis of intervention planning and monitoring. Developing items appropriate for younger ages to extend the ability to use a naturalistic probe measure for preschool and young school-age children would also extend the utility of this measure.

In addition, examining the instrument in other populations at risk for social difficulties, such as children with learning disabilities, attention deficit disorders, or nonautistic social communication impairments, would provide broader support for the use of a probe measure for documenting pragmatic deficits. Studies of response to intervention that attempt to use dynamic assessments like the Y*i*PP to document change will also contribute to understanding the potential usefulness of these kinds of procedures and will lay the basis for broader availability of dynamic assessment for pragmatic language skills.

Acknowledgments

Preparation of this article was supported by Research Grant P0-HD03008 funded by the National Institute of Child Health and Human Development (NICHD)—Autism and Related Disorders: Development and Outcome. We would like to thank the families who participated in this study, in addition to Rebecca Loomis, who assisted with the behavioral coding.

References

Adams, C., Green, J., Gilchrist, A., & Cox, A. (2002). Conversational behaviour of children with Asperger syndrome and conduct disorder. *Journal of Child Psychology and Psychiatry, 43,* 679–690.

Adams, C., & Lloyd, J. (2005). Elicited and spontaneous communicative functions and stability of conversational measures with children who have pragmatic language impairments. *International Journal of Language and Communication Disorders, 40,* 333–348.

Adams, C., Lockton, E., Freed, J., Gaile, J., Earl, G., McBean, K., . . . Law, J. (2012). The Social Communication Intervention Project: A randomized controlled trial of the effectiveness of speech and language therapy for school-age children who have pragmatic and social communication problems with or without autism spectrum disorder. *International Journal of Language & Communication Disorders, 43,* 233–244.

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Speech-Language-Hearing Association. (2014). *Social language use (pragmatics).* Retrieved from http://www.asha.org/public/speech/development/Pragmatics

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: Cognitive or affective? *Journal of Autism and Developmental Disorders, 18,* 379–402.

Bellini, S., Gardner, L., & Markoff, K. (2014). Social skill interventions. In F. Volkmar, R. Paul, S. Rogers, & K. Pelphrey (Eds.), *Handbook of autism spectrum disorder* (5th ed., pp. 887–906). New York, NY: Wiley.

Bishop, D. (2006). *Children's Communication Checklist–Second Edition*. London, England: Psychological Corporation.

Bishop, D., & Adams, C. (1989). Conversational characteristics of children with semantic–pragmatic disorders: What features lead to a judgment of inappropriacy? *British Journal of Disorders of Communication, 24,* 241–261.

Brinton, B., & Fujiki, M. (1993). Clinical forum: Language and social skills in the school-age population. *Language, Speech, and Hearing Services in Schools, 24,* 194–198.

Brinton, B., Robinson, L., & Fujiki, M. (2004). Description of a program for social language intervention: "If you can have a conversation, you can have a relationship." *Language, Speech, and Hearing Services in Schools, 35,* 283–290.

Bryan, T., Burstein, K., & Ergul, C. (2004). The social-emotional side of learning disabilities: A science-based presentation of the state of the art. *Learning Disability Quarterly, 27,* 45–51.

Chapman, R. (1981). Exploring children's communicative intents. In J. Miller (Ed.), *Assessing language production in children* (pp. 111–138). Boston, MA: Allyn & Bacon.

Chuba, H., Paul, R., Klin, A., & Volkmar, F. (2003, November). *Assessing pragmatic skills in individuals with autism spectrum disorders*. Presentation at the National Convention of the American Speech-Language-Hearing Association, Chicago, IL.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., . . . Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview–Revised. *Journal of Autism and Developmental Disorders, 33,* 427–433.

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

Creaghead, N. (1984). Strategies for evaluating and targeting pragmatic behaviors in young children. *Seminars in Speech and Language, 5,* 241–252.

Dawson, M., Mottron, L., & Gernsbacher, M. A. (2008). Learning in autism. In J. H. Byrne & H. Roediger (Eds.), *Learning and memory: A comprehensive reference: Cognitive psychology* (pp. 759–772). New York, NY: Elsevier.

Dewey, M., & Everard, M. (1974). The near normal autistic adolescent. *Journal of Autism and Childhood Schizophrenia, 4,* 348–356.

Elliot, C. (2007). *Differential Ability Scales–II: School Age Edition*. San Antonio, TX: Pearson.

Feldman, H., Dale, P., Campbell, T., Colborn, D., Kurs-Lasky, M., Rockette, H., & Paradise, J. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76,* 856–868.

Fombonne, E. (2003). Epidemiological surveys of autism and other pervasive developmental disorders: An update. *Journal of Autism and Developmental Disorders, 33,* 365–382.

Gadow, K., & Sprafkin, J. (2002). *Childhood and Adolescent Symptom Inventory 4: Screening and norms manual*. Stony Brook, NY: Checkmate Plus.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference: 11.0 update*. (4th ed.). Boston, MA: Allyn & Bacon.

Gillberg, C., & Ehlers, S. (1998). High-functioning people with autism and Asperger syndrome: A literature review in E. Schopler & G. M. Mesibov (Eds.), *Asperger syndrome or high functioning autism?* (pp. 79–106). New York, NY: Plenum.

Gutiérrez-Clellen, V., & Peña, E. (2001). Dynamic assessment of diverse children: A tutorial. *Language, Speech, and Hearing Services in Schools, 32,* 212–224.

Kim, S., Paul, R., Tager-Flusberg, H., & Lord, C. (2014). Communication in ASD. In F. Volkmar, R. Paul, S. Rogers, & K. Pelphrey (Eds.), *Handbook of autism spectrum disorders* (4th ed., pp. 230–262). New York, NY: Wiley.

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society B, 358,* 345–360.

Landa, R. (2000). Social language use in Asperger syndrome and high-functioning autism. In A. Klin, F. Volkmar, & S. Sparrow (Eds.), *Asperger syndrome* (pp. 125–155). New York, NY: Guilford Press.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

Larson, V., & McKinley, N. (1995). *Language disorders in older students: Preadolescents and adolescents*. Eau Claire, WI: Thinking Publications.

Lidz, C., & Peña, L. (1996). Dynamic assessment: The model, its relevance as a nonbiased approach, and its application to Latino American preschool children. *Language, Speech, and Hearing Services in Schools, 27,* 367–372.

Loomis, R., Schoen, E., & Paul, R. (2011, May). *Assessing pragmatic language: The Yale* in vivo *Pragmatic Protocol*. Poster presented at the International Meeting for Autism Research, Toronto, Ontario, Canada.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2001). *Autism Diagnostic Observation Schedule*. Los Angeles, CA: Western Psychological Services.

Lyons, M., Simmons, E. S., & Paul, R. (2014). Prosodic development in middle childhood and adolescence in high functioning autism. *Autism Research, 7*(2), 181–196.

Miles, S., Paul, R., Klin, A., Volkmar, F., & Cohen, D. (2001, November). *Pragmatic assessment of conversation in high-functioning autism*. Poster presented at the National Convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Norbury, C. (2014). Practitioner review: Social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *Journal of Child Psychology and Psychiatry, 55,* 204–216.

O'Neil, D. (2009). *Language Use Inventory*. Ontario, Canada: Knowledge in Development.

Paul, R. (2005). Assessing communication in autism. In F. Volkmar, A. Klin, R. Paul, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd ed., pp. 799–816). New York, NY: Wiley.

Paul, R., & Fahim, D. (2014). Assessing communication in autism. In F. Volkmar, R. Paul, S. Rogers, & K. Pelphrey (Eds.), *Handbook of autism spectrum disorders* (4th ed., pp. 673–694). New York, NY: Wiley.

Paul, R., Landa, R., & Simmons, E. (2014). Assessing and treating communication. In J. McPartland, A. Klin, & F. Volkmar (Eds.), *Asperger syndrome: Assessing and treating high-functioning autism spectrum disorder* (2nd ed., pp. 103–142). New York, NY: Guilford Press.

Paul, R., & Norbury, C. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (4th ed.). St. Louis, MO: Mosby.

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.

Prutting, C., & Kirchner, D. (1983). Applied pragmatics. In T. M. Gallagher & C. A. Prutting (Eds.), *Pragmatic assessment and intervention issues in language* (pp. 29–64). San Diego, CA: College-Hill Press.

Redman, S., & Rice, M. (2002). Stability of behavioral ratings in children with SLI. *Journal of Speech, Language, and Hearing Research, 45,* 190–201.

Reichow, B., Salamak, S., Paul, R., Volkmar, F., & Klin, A. (2008). Pragmatic assessment in autism spectrum disorders: A comparison of a standard measure with parent report. *Communication Disorders Quarterly, 29,* 169–176.

Rice, M., & Buhr, J. (1992). Specific-language-impaired children's quick incidental learning of words: The effect of a pause. *Journal of Speech and Hearing Research, 35,* 1040–1049.

Rice, M., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1412–1431.

Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.

Rutter, M., LeCouteur, A., & Lord, C. (2003). *Autism Diagnostic Interview, Revised*. Los Angeles, CA: Western Psychological Services.

Schoen, E., & Paul, R. (2009, April). *Assessing pragmatic language skills in high-functioning autism spectrum disorders*. Poster presented at the Society for Research in Child Development, Denver, CO.

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition*. San Antonio, TX: Pearson.

Siegel, D., Minshew, N., & Goldstein, G. (1996). Wechsler IQ profiles in diagnosis of high-functioning autism. *Journal of Autism and Developmental Disorders, 26,* 389–406.

Swanson, H., & Lussier, C. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research, 71,* 321–363.

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders, 11,* 45–56.

Volkmar, F. (1987). Social development. In D. Cohen, A. Donnellan, & R. Paul (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 41–60). New York, NY: Wiley.

Volkmar, F., Klin, A., & Rutter, M. (2005). Issues in classification of autism and related conditions. In F. Volkmar, A. Klin, R. Paul, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd ed., pp. 5–41). New York, NY: Wiley.

Volkmar, F., Lord, C., Bailey, A., Schultz, R., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 45,* 135–170.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Pearson.

Weiner, J. (2002). Friendship and social adjustment of children with learning disabilities. In M. L. Donahue & B. Wong (Eds.), *The social dimensions of learning disabilities: Essays in honor of Tanis Bryan* (pp. 93–114). New York, NY: Routledge.

Weiss, A., Paul, R., & Norbury, C. (2006, November). *Pragmatic language impairments*. Seminar presented at National Convention of the American Speech-Language-Hearing Association, Miami, FL.

Young, E., Diehl, J., Morris, D., Hyman, S., & Bennetto, L. (2005). The use of two language tests to identify pragmatic language problems in children with autism spectrum disorders. *Language, Speech, and Hearing Services in Schools, 36,* 62–72.