JSLHR

Research Article

Story Goodness in Adolescents With Autism Spectrum Disorder (ASD) and in Optimal Outcomes From ASD

Allison R. Canfield,[a] Inge-Marie Eigsti,[a]
Ashley de Marchena,[b] and Deborah Fein[a]

Purpose: This study examined narrative quality of adolescents with autism spectrum disorder (ASD) using a well-studied "story goodness" coding system.
Method: Narrative samples were analyzed for distinct aspects of story goodness and rated by naïve readers on dimensions of story goodness, accuracy, cohesiveness, and oddness. Adolescents with high-functioning ASD were compared with adolescents with typical development (TD; n = 15 per group). A second study compared narratives from adolescents across three groups: ASD, TD, and youths with "optimal outcomes," who were diagnosed with ASD early in development but no longer meet criteria for ASD and have typical behavioral functioning.
Results: In both studies, the ASD group's narratives had lower composite quality scores compared with peers with typical development. In Study 2, narratives from the optimal outcomes group were intermediate in scores and did not differ significantly from those of either other group. However, naïve raters were able to detect qualitative narrative differences across groups.
Conclusions: Findings indicate that pragmatic deficits in ASD are salient and could have clinical relevance. Furthermore, results indicate subtle differences in pragmatic language skills for individuals with optimal outcomes despite otherwise typical language skills in other domains. These results highlight the need for clinical interventions tailored to the specific deficits of these populations.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social interaction and communication and the presence of restricted and repetitive behaviors and interests (American Psychiatric Association [APA], 2013). Pragmatic language, or socially appropriate communication, lies at the intersection of the social and communication deficits in ASD, which can thus interfere with social interactions and affect academic and occupational functioning. Deficits in pragmatic language are observed even in children who no longer meet diagnostic criteria for ASD and are thought to have achieved "optimal outcomes" (OO; Kelley, Paul, Fein, & Naigles, 2006). Pragmatic language deficits in ASD warrant further investigation to illuminate underlying mechanisms; this is a primary objective of the current article.

Pragmatic language includes spoken qualities, such as word choice and content, turn-taking, and prosody, and nonverbal communicative behaviors, such as eye contact, gestures, and facial expressions (Eigsti, de Marchena, Schuh, & Kelley, 2011). The presentation of general language abilities in ASD can vary widely, encompassing individuals who are functionally nonverbal (Lord & Paul, 1997) and those whose language abilities are within typical limits (Landa, 2000). Children with ASD who are verbal may show delayed acquisition of vocabulary and grammar (Eigsti & Bennetto, 2009; Eigsti, Bennetto, & Dadlani, 2007) but may have average or above-average scores on standardized tests (Kjelgaard & Tager-Flusberg, 2001; Tager-Flusberg, Joseph, & Folstein, 2001). Pragmatic language impairments appear more resistant to remediation or improvement than basic language (Landa, 2000; Lord & Paul, 1997) and are considered to be a universal deficit across the autism spectrum.

Although there are many standardized assessments of language structure, assessments of pragmatic language deficits are limited (Young, Diehl, Morris, Hyman, & Bennetto, 2005), because much of social communication

[a]University of Connecticut, Storrs
[b]Center for Autism Research, Intellectual and Developmental Disabilities Research Center, The Children's Hospital of Philadelphia, PA
Correspondence to Inge-Marie Eigsti: inge-marie.eigsti@uconn.edu

Editor: Rhea Paul
Associate Editor: Joanne Volden
Received January 23, 2015
Revision received June 3, 2015
Accepted October 14, 2015
DOI: 10.1044/2015_JSLHR-L-15-0022

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

is context dependent and requires flexibility, both of which are difficult to evaluate in standardized assessments. Indeed, the highly structured format of standardized tests is likely why these tests are often a strength in individuals with ASD (Kelley et al., 2006). The ideal approach for characterizing pragmatic language skills is an ongoing debate. Although some studies report that standardized measures, such as the Test of Pragmatic Language (Phelps-Terasaki & Phelps-Gunn, 1992), or parent-report measures, such as the Children's Communication Checklist-2 (Bishop, 2006), adequately capture poor performance in ASD (Volden & Phillips, 2010; Young et al., 2005), other studies suggest that parent report of language abilities is less accurate in school-aged children and adolescents (Dale, 1996; Hauerwas & Stone, 2000). Furthermore, the Test of Pragmatic Language is only normed up to age 14 years; it and other standardized measures have proven limited in their ability to detect subtle pragmatic deficits (Volden & Phillips, 2010). These limitations highlight the need for a sensitive assessment tool that permits direct examiner observation of pragmatic language abilities in order to best evaluate subtle pragmatic strengths and weaknesses in older children and adolescents.

Narrative as a Measure of Pragmatic Language

Given limitations in the standardized assessments of pragmatic abilities, many researchers and clinicians have turned to narrative elicitation. In such studies, participants use a structured narrative prompt, such as a picture book, to produce a story that is subsequently transcribed and coded. The production of a spoken narrative is a complex task, tapping semantic, grammatical, and pragmatic abilities, as well as planning and organization (Diehl, Bennetto, & Young, 2006), among other skills. Narrative production may be more complex and demanding than conversation (Abbeduto, Benson, Short, & Dolish, 1995; Botting, 2002), because conversational partners provide priming and structure for subsequent turns in a dialogue (Garrod & Pickering, 2004). In narrations, syntactic complexity, lexical diversity, and narrative length all increase with increasing language skill (Botting, 2002). Narrative ability relates to academic and social success in both children with typical development and children with language and learning disabilities (Boudreau, 2008; Fox & Wright, 1997); furthermore, narratives have high ecological validity.

Narratives in ASD

Narrative studies have examined semantic language abilities in ASD, evaluating the use of mental state verbs, or descriptions of a character's thoughts or feelings (Capps, Losh, & Thurber, 2000), which relate to theory of mind abilities. Although some narration studies reported decreased mental state language in ASD (Losh & Capps, 2003), others report no such differences (Beaumont & Newcombe, 2006; Capps et al., 2000; Norbury & Bishop, 2003; Tager-Flusberg & Sullivan, 1995).

Narrative elicitation has also been used to examine narrative length, structure, and cohesion in individuals with ASD. Studies comparing children with ASD and those with typical development (TD), with groups matched on language and cognitive abilities, report minimal quantitative differences in narrative length and syntactic complexity but deficits in organization and cohesiveness (Diehl et al., 2006; Tager-Flusberg & Sullivan, 1995). Narratives in ASD tend to lack coherent global organization (Baron-Cohen, Leslie, & Frith, 1986; Loveland, McEvoy, Tunali, & Kelley, 1990; Suh et al., 2014), with fewer causal connections between story events (cohesion) and reduced communication of story "gist" (Diehl et al., 2006).

In addition to story organization, an important metric of narrative quality is *completeness*. Suh et al. (2014) reported that children with high-functioning autism (HFA; IQ > 80) provided less complete narratives than children with TD, where groups did not differ in age, gender, or verbal IQ. Norbury, Gemmell, and Paul (2014) reported that children with ASD ages 6 to 15 years produced narratives with more pragmatic errors, less cohesive narrative macrostructure, and omission of significant story elements. This completeness deficit persists into late adolescence and adulthood; adults and adolescents with HFA produced narratives of similar length (i.e., number of independent clauses) but with fewer complete story episodes, compared with a TD group (Landa, Martin, Minshew, & Goldstein, 1995).

Optimal Outcomes From ASD

ASD has historically been seen as a lifelong disorder; however, recent studies have explored the phenomenon of "optimal outcomes" from ASD, in which individuals with a well-documented early ASD diagnosis no longer meet the diagnostic criteria. Fein et al. (2013) described a sample of 34 youth with OO who, in addition to losing their ASD diagnosis, were indistinguishable from peers with TD on standardized measures of language, socialization, and adaptive skills. Studies have described normalized performance by OO groups on standardized assessments of language abilities (Tyson et al., 2014), with better performance in comparison to peers with HFA, who continued to show deficits.

Relative to these findings of fully normative language abilities, one study of younger children with OO (mean age of 8 years) reported subtle continued deficits on some pragmatic language measures (i.e., theory of mind, narrative performance, and use of mental state verbs). Participants with OO had narrative deficits, including fewer causal connections, less clear referents, and less discussion of the goals and motivations, compared with peers with TD (Kelley et al., 2006). A subsequent study of participants with OO ages 9 to 15 years, including some of the same children from the 2006 study, reported higher rates of idiosyncratic language (e.g., highly specific or unusual references to topics outside the scope of the story) and speech disfluencies in the OO group (Suh et al., 2014). Pragmatic difficulties may be particularly resistant to remediation, as

the children with OO in this study were functioning in the average and above-average ranges on standardized language measures.

Story Grammar

One methodological tool used in pragmatic language research is narrative elicitation. It has been used to examine subtle communication deficits and how story organization breaks down following traumatic brain injury (Hartley & Jensen, 1991; Mozeiko, Lê, Coelho, Krueger, & Grafman, 2011). Some studies use *story grammar*, which examines the organization of story components and the logical relationships between characters and story events, to assess narrative quality (Stein & Glenn, 1979; Thorndyke & Yekovich, 1980). A story grammar quantifies narrative organization into prototypical story structures, such as initiating actions toward character goals, attempts to solve problems, and the consequences of those attempts (Liles, Coelho, Duffy, & Zalagens, 1989; Liles, Duffy, Merritt, & Purcell, 1995). Poorly organized narratives provide clues to cognitive, semantic, and pragmatic impairments, and story grammar provides a reliable tool to evaluate such narratives (Lê, Coelho, Mozeiko, & Grafman, 2011).

One story grammar metric (*story goodness*) includes measures of narrative completeness and narrative organization. After viewing a wordless cartoon story, participants tell the story. This measure distinguishes the narratives of adults with and without brain injury; the former produce less complete and organized narratives (Lê et al., 2011), and the combined metric better characterizes performance than either component alone. Although previous studies of ASD have included elements of story grammar (Young et al., 2005), none have yet used the informative metric of story goodness. Narrative analysis using a story-goodness coding system will permit comparisons between people with ASD and other clinical groups, and may reveal subtle pragmatic language deficits even in a sample of high-functioning individuals with intact language abilities.

Current Study

The objective of the present study was to evaluate pragmatic language abilities via narrative quality in two sets of participants: Study 1 involved adolescents with HFA and TD, and Study 2 involved adolescents with OO, HFA, and TD. We evaluated both specific categories of pragmatic language impairment, as well as global pragmatic language abilities. We assessed narrative quality in several ways, including story-goodness coding (Lê et al., 2011) and a novel measure, the *story composite* score, both of which combine specific coding categories in order to examine global narrative quality. On the basis of previous research, we predicted that, compared with a TD comparison group, stories by adolescents with HFA would be of similar length but be less complete. In contrast, we predicted that individuals with OO would not differ from their TD peers. There have been mixed findings regarding narrative organization in ASD; with this measure of story grammar, we predicted that performance in this domain would be similar across groups, similar to the story-grammar findings of Young et al. (2005). We predicted that youths with HFA would receive significantly lower story composite scores than the TD group, with the OO group's performance falling between the two, given the pragmatic weaknesses previously found in OO (Suh et al., 2014). To evaluate the ecological validity of story goodness, we acquired ratings of narrative quality from naïve raters, for which we predicted that participants with HFA would receive lower ratings for story goodness and cohesiveness, but not for story accuracy and oddness. Overall, we predicted that the narratives from the HFA groups would have lower story-goodness scores and that narratives from the OO group would not differ from those from a TD comparison group.

Study 1

Method

Participants

Participants were 20 adolescents with HFA and 16 with TD; the groups did not differ on chronological age, gender, or full-scale IQ. Participants were recruited via fliers in the community, by word of mouth, and through previous studies. For inclusion, participants were required to have an IQ above 80 and be native speakers of English. Participants with comorbid learning or psychiatric disorders were not excluded, to allow for normative heterogeneity in the sample. Participants were excluded from the TD group if they had a history of neurological problems or any first-degree relatives with ASD. Within the HFA group, one participant was excluded due to equipment failure, one due to a failure to confirm HFA diagnosis, and three due to having an IQ below 80. One participant was excluded from the TD group for having a history of a mild neurological impairment. Altogether, five participants from the HFA group and one from the TD group were excluded, and the final sample of 30 incorporated 15 adolescents each with HFA and TD. Demographic information is shown in Table 1.

Diagnoses were verified for the HFA participants and ruled out for the TD participants by research-reliable clinicians using the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2002), the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003), and criteria from the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; APA, 1994). Prior to testing, written consent was obtained from parents and assent was obtained from participants. This research was approved by the University of Connecticut Institutional Review Board.

Measures

The Stanford–Binet Intelligence Scales, Fifth Edition (Roid, 2003), are a measure of intelligence and cognitive abilities. Participants completed the Matrices and Vocabulary

Table 1. Demographic information for participants with high-functioning autism (HFA) and typical development (TD): Study 1.

Characteristic	HFA	TD	χ^2 or F	p	η_p^2
N (M;F)	15 (13;2)	15 (14;1)	0.37	.54	
Chronological age (years)	15.0 (1.5), 12.4 to 17.3	15.0 (1.5), 12.8 to 17.6	0.01	.93	< .01
SB5 NVIQ	10 (1.6), 7 to 12	11 (2.0), 6 to 13	2.68	.11	.09
SB5 VIQ	11 (2.6), 6 to 16	10 (1.6), 7 to 13	1.63	.21	.06
SB5 FSIQ	103 (9.9), 85 to 118	103 (8.9), 82 to 115	0.003	.95	< .01
PPVT-3	110 (12.8), 78 to 127	117 (8.0), 105 to 137	3.37	.08	.11
ADOS Communication	2.7 (1.0), 1 to 4	0.3 (1.0), 0 to 1	68.60	< .001	.71
ADOS Social	7.3 (1.9), 2 to 10	0.3 (0.6), 0 to 2	184.19	< .001	.87
SCQ (total score)	20.6 (6.8), 10 to 29	1.9 (2.9), 0 to 9	88.09	< .001	.78

Note. Data are shown as *M* (*SD*), range. SB5 = Stanford–Binet Intelligence Scales, Fifth Edition; NVIQ = nonverbal IQ; VIQ = verbal IQ; FSIQ = full-scale IQ; PPVT-3, Peabody Picture Vocabulary Test–III; ADOS = Autism Diagnostic Observation Schedule; SCQ = Social Communication Questionnaire.

subtests to obtain estimates of nonverbal (NVIQ) and verbal (VIQ) cognitive functioning, from which a full-scale IQ (FSIQ) score was calculated. The Peabody Picture Vocabulary Test–III (PPVT-3; Dunn & Dunn, 1997) is a standardized measure of receptive vocabulary. The Social Communication Questionnaire is a standardized parent-report measure that identifies symptoms of ASD. Parents completed the Lifetime version. Data from 27 parents are reported, because three parents (two from the HFA group, one from the TD group) did not return the questionnaire. The ADOS is a semistructured ASD assessment. The Monkey Cartoon task of the ADOS, in which participants produce a narrative on the basis of a standard stimulus, provided the narrative measure for Study 1.

Procedure

Participants were tested in a quiet room at their home, their school, or the University of Connecticut. The measures included in this study were collected as part of a larger study of language in ASD. Testing took approximately 4 hr over one or two sessions.

Stimuli for the narrative task were six black-and-white drawings depicting a story about two monkeys. This story incorporated the following events:

1. Monkey A is up in a tree picking coconuts.
2. Monkey B is leaning against a tree and a coconut lands in front of it; Monkey B appears surprised.
3. Monkey B picks up the coconut.
4. Monkey A looks down to where the coconut was and appears confused. Monkey B is running away with the coconut.
5. Monkey A drops a second coconut from the tree; Monkey B runs to retrieve it.
6. Monkey B leans over to take the new coconut. Monkey A throws a coconut and hits Monkey B on the head.

The examiner told participants that they would be asked to tell the story shown in the cards and then laid out one picture card at a time. Participants were encouraged to examine each card for as long as they liked. The cards were removed, and participants were instructed to stand up and tell the story. The narrations were recorded on digital video.

Speech Transcription

Narratives were transcribed from the video using Computerized Language Analysis (MacWhinney, 2000). All words and disfluencies were transcribed, and speech was broken into minimal terminal units, known as T-units (roughly equivalent to a spoken sentence), which consist of a main clause and any attached or embedded dependent clauses (Hunt, 1965).

Coding

The completeness and organization of narratives was coded as described by Lê et al. (2011). Story completeness was operationalized as the number of core story events, out of a possible six, described in the story; the *story-completeness score* was the number of core story events expressed, with a maximum score of 6.

Story grammar or story organization was coded following Lê et al. (2011) and Merritt and Liles (1987). Each T-unit was evaluated as falling into one of three mutually exclusive categories: (a) *initiating events*, which prompt a character to act (e.g., *and he dropped one [a coconut] on the ground*); (b) *attempts*, describing actions related to the initiating event (e.g., *the monkey below him took it and ran away with it*); and (c) *direct consequences*, the results of attempts or the attainment or nonattainment of a goal (e.g., *and when the monkey below him was about to run away, the monkey in the tree threw a coconut*). The *story-grammar score* was calculated as the sum of T-units that fell into any of these three categories, divided by the total number of T-units in the narrative (Lê et al., 2011). A sample coded transcript is provided in the Appendix.

In addition to these three categories (initiating events, attempts, and direct consequences), there were three additional possible categories. As described in Merritt and Liles (1987), we coded the following: (a) *setting*, which captured

information about a character's personality, habits, or location (e.g., *there was a monkey up in a tree*), and (b) *internal response*, which captured mental state descriptions of the thoughts and feelings of characters, similar to previous studies of mental state word use in ASD (e.g., *he gets all excited that he found the coconut*). Following the findings of Suh et al. (2014) that children with HFA were 21 times more likely to use idiosyncratic language during narrative production, we also included an exploratory analysis with (c) *added/invented details*, which captured off-topic or fictitious details (e.g., *he wants to make a coconut banana cream pie that everyone loves*). Setting and internal response were coded according to criteria outlined by Merritt and Liles (1987); criteria for added/invented detail are described in the Appendix.

Because our aim was to characterize global pragmatic abilities, in addition to specific areas of deficit, we calculated a *story composite score* by combining scores from six coding categories—completeness, initiating events, attempts, direct consequences, setting, and internal response—each of which provides a measure of specific narrative qualities and contributes to the global quality of a narrative. Z-scores were calculated for each of the six variables, and the story composite score was calculated as an average of these z-scores. Added/invented detail and total T-units were not included, because extraneous detail and excessive length detract from narrative quality.

Reliability

To assess speech-transcription reliability, two independent coders transcribed nine narratives (four from the HFA group, five from the TD group). The intraclass correlation coefficient (ICC) for transcription was .95. The first author coded each of the 30 narratives without knowing diagnoses. A trained research assistant, who also did not know the diagnoses, coded 36% of the transcripts (n = 11; five from the ASD group and six from the TD group) for interrater reliability. Analyses indicated good interrater reliability for story completeness (ICC = .61) and excellent reliability for story grammar (ICC = .75), setting (ICC = .91), internal response (ICC = .86), and added/invented detail (ICC = .95; Cicchetti, 1994).

Story-Quality Ratings

Thirty-nine University of Connecticut undergraduate students rated narrative quality in return for extra course credit. Raters were over 18 years of age and were native speakers of English. The raters were unaware of the study's hypotheses. They first viewed the monkey cartoons and told the story, which was recorded; six raters were excluded because their own narratives suggested poor comprehension of the story. One rater was excluded due to a technical error. After narrating the story themselves, raters read 10 transcribed narratives (presented in random order, with five HFA and five TD narratives for each rater). Disfluencies (i.e., "uh," "um," "err") and false starts (i.e., "the monkey came ba-ba-back") were removed from transcriptions to prevent them from influencing ratings.

After reading a written transcription of a narrative, raters answered four questions about each narrative. Rating questions, presented in order, were the following: "How good a story is this?" (i.e., goodness); "How well were you able to follow this story?" (i.e., cohesiveness); "How well does this story reflect the actual story in the cards?" (i.e., accuracy); and "How odd/unusual did you find this story?" (i.e., oddness). Ratings were provided using a 1–5 Likert scale (e.g., for goodness, a score of 1 corresponded to a rating of *not good* and a score of 5 corresponded to a score of *very good*). Narratives, rating questions, and responses were presented and collected using SuperLab.

Results

To examine the narrative quality of adolescents with HFA, analyses compared several measures of story goodness as well as more specific components of narrative quality, from participants with both HFA and TD.[1] Dependent variables were examined for deviations from the assumptions of normality and sphericity and were found to be normally distributed. Data were checked for outliers (scores > 2 *SD*s from the group mean). No data were excluded. Effect sizes were calculated as partial eta squared (η_p^2), which refers to the proportion of variance attributable to a given effect after partialing out nonerror sources of variance (Cohen, 1988). Data and statistical analyses are provided in Table 2.

Narrative Analysis

Story composite score. To capture overall narrative quality, a story composite score was created summing across six coding categories. The story composite score for the HFA group was significantly lower than that for the TD group, $F(1, 28) = 10.24$, $p = .003$, $\eta_p^2 = .27$.

Story completeness. Story completeness was quantified by the number of essential plot events a participant included in the narrative (range: 0 to 6). The mean completeness score for the HFA group was lower than that for the TD group, though the difference did not reach significance, $F(1, 28) = 3.54$, $p = .07$, $\eta_p^2 = .11$.

Story grammar. Story grammar was quantified by the ratio of initiating events, attempts, and direct consequences to the total number of T-units. The story-grammar ratio did not differ between groups, $F(1, 28) = 0.07$, $p = .80$, $\eta_p^2 = .002$. Figure 1 presents story goodness by plotting story completeness against story grammar, with the best scores on both leading to placement in the upper right quadrant.

Additional coding categories. Setting, internal response, added/invented detail, and total T-units were analyzed separately. Data are shown in Table 2. Although the scores for the HFA group were numerically higher for

[1]To determine whether general language ability contributed to any differences, analyses of narrative measures were repeated with VIQ as a covariate; results were unchanged. Given arguments that including covariates that also capture the variable of interest may obscure disorder-specific patterns (Dennis et al., 2009), the analyses are reported here without covariates.

Table 2. Story-goodness results: Study 1.

Measure	HFA	TD	*F*	*p*	η_p^2
Story composite[a]	–0.29 (0.54), –1.35 to 0.95	0.29 (0.46), –0.26 to 1.12	10.24	.003	.27
Story completeness[b]	4.4 (1.5), 2 to 6	5.2 (0.80), 4 to 6	3.54	.07	.11
Story grammar[c]	.60 (.23), .17 to 1.0	.62 (.20), .23 to .60	0.07	.80	.002
Coding categories[d]					
Setting	0.40 (0.83), 0 to 3	1.07 (1.33), 0 to 5	2.70	.11	.09
Internal response	0.93 (0.88), 0 to 3	1.47 (1.41), 0 to 4	1.55	.22	.05
Added/invented detail	2.13 (3.50), 0 to 13	1.40 (1.64), 0 to 5	0.54	.47	.02
Total T-units	9.73 (4.13), 5 to 19	11.8 (4.50), 7 to 22	1.72	.20	.06

Note. Data are shown as *M* (*SD*), range. HFA = high-functioning autism; TD = typical development.

[a]Average z-scores. [b]Average score out of 6 total. [c]Average proportion (i.e., sum of T-units that fell into the categories of initiating event, attempt, and direct consequences divided by the total number of T-units in the narrative). [d]Average frequency.

added/invented detail and numerically lower for all other categories, group differences were not significant.

Coding results indicated that whereas the HFA group narrations were less coherent, organized and complete than those of the TD group, there was no single quality that differed by group. Instead, the constellation of a variety of factors drove the group differences.

Story-Quality Ratings

College undergraduates, unaware of the study hypotheses, rated each of 10 narratives on four dimensions. Interrater reliability was calculated using one-way random average ICCs. Results indicated that interrater reliability was excellent across all four categories (goodness: ICC = .85; cohesiveness: ICC = .83; accuracy: ICC = .84; and oddness: ICC = .83).

A repeated measures analysis of variance revealed a significant main effect of Characteristic Rating, $F(3, 84) = 3.73$, $p = .01$, $\eta_p^2 = .12$, a significant main effect of Group, $F(1, 28) = 1355.15$, $p < .001$, $\eta_p^2 = .98$, and no Dimension Rating × Group interaction, $F(3, 84) = 1.59$, $p = .20$, $\eta_p^2 = .05$. Follow-up univariate analyses were conducted; data are shown in Table 3. Cohesiveness ratings differed by group. Goodness ratings approached but did not reach significance. Ratings were lower for the HFA group than the TD group in both cases. Correlations between goodness ratings, story composite scores, and age were examined, but none were significant.

Discussion

This study evaluated pragmatic language abilities and narrative quality in adolescents with HFA and TD using a story-goodness coding system and narrative-quality ratings. These measures may reveal even very subtle pragmatic language differences in individuals with ASD who are comparable to their peers with TD in structural and vocabulary language domains.

A composite measure of story quality showed a clear group difference, suggesting a deficit in pragmatic language abilities in this sample of youth with HFA. No specific measures of story goodness carried this difference, suggesting that stories may be of poor quality along multiple dimensions although not grossly atypical in any single dimension.

The results of this study indicate that adolescents with HFA tell stories that are less complete than their peers

Figure 1. Study 1 story goodness. Story goodness is a function of story grammar and story completeness. Lines dividing the figures into quadrants are set at 1 *SD* below the group mean for typical development. The upper right quadrant is where scores for the best stories fall, with high scores for both story grammar and completeness. The narrative performance of the group with high-functioning autism spectrum disorder appeared more variable than that of the group with typical development, especially on the completeness measure.

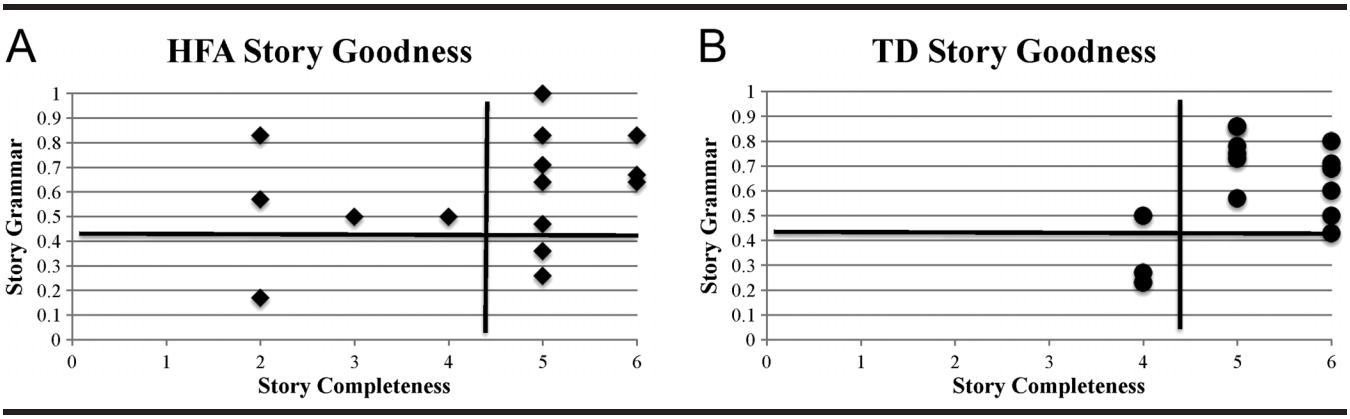

Table 3. Narrative-quality ratings: Study 1.

Dimension	HFA	TD	*F*	*p*	η_p^2
Goodness	2.46 (0.84), 1.3 to 4.0	3.08 (0.88), 1.4 to 4.5	3.94	.06	.12
Cohesiveness	2.90 (0.95), 1.5 to 4.7	3.54 (0.71), 2.3 to 4.7	4.49	.05	.14
Accuracy	3.00 (0.85), 1.6 to 4.5	3.28 (0.83), 2.3 to 4.7	0.84	.37	.03
Oddness	2.70 (0.85), 1.4 to 4.3	2.55 (0.66), 1.6 to 3.8	0.3	.59	.01

Note. Data are shown as *M* (*SD*), range. Scores are represented as average ratings on a 5-point scale. HFA = high-functioning autism; TD = typical development.

with TD, consistent with prior research (Landa et al., 1995; Suh et al., 2014). Despite being less complete (e.g., fewer core plot events were expressed), the HFA group's narratives were no shorter, suggesting that the HFA group was less efficient than the TD group in storytelling. Consistent with predictions, the two groups were virtually indistinguishable on story grammar, which suggests that the HFA group framed their narratives around prototypical story events in a manner similar to the TD group, indicating this as a strength. The findings suggested that individuals with HFA do not struggle with organization during the production of a short narrative from memory the way they do when producing a longer story while narrating a picture book (Diehl et al., 2006; Suh et al., 2014). Intact story-grammar ability may contribute to the impaired story-completeness score in the HFA group, in that greater cognitive resources were dedicated to appropriately structuring the narrative at the expense of recalling key story events.

There were no group differences in the use of mental state descriptions, which have often been studied in ASD (Beaumont & Newcombe, 2006; Capps et al., 2000; Norbury & Bishop, 2003; Tager-Flusberg & Sullivan, 1995). In addition, whereas HFA narratives are reported to include more idiosyncratic language (Suh et al., 2014), there were no group differences in this study in the presence of added or invented details. The short, simplistic plot of the narrative picture stimuli may account for this lack of a quantitative difference between the groups on this measure, because such a straightforward plot does not allow for as much creativity. Our study may also have been underpowered to detect these differences; both effects were in the predicted direction, though neither reached significance.

Naïve raters were sensitive to differences in narrative quality on the basis of questions regarding goodness and cohesiveness, consistent with other studies (de Marchena & Eigsti, 2016). It is particularly interesting that raters detected group differences just by reading the transcribed narrations, in the absence of speech characteristics such as disfluency, prosody, or appearance. This suggests that differences in narrative quality are quite salient and likely play a role in the daily social interactions of individuals with ASD.

In summary, the findings of Study 1 are consistent with previous narrative research in ASD. First, a broad measure of narrative quality, created from summed scores across more specific narrative components (i.e., story composite score), showed that the HFA group's narratives were of significantly lower quality. These data suggest that the narratives of the HFA group were less complete, yet similar in length, indicating that the HFA group was less efficient in storytelling. During the production of this short narrative, the HFA and TD groups were similar in their story structure as measured by story grammar. These coded data were consistent with the findings of lower ratings of goodness and cohesiveness in HFA. These results indicate that it is the combination of multiple aspects of narrative production that contributes to an overall poorer narrative quality in HFA, rather than specific aspects of narrative performance in isolation.

Study 2

Method

Participants

Fifteen children and adolescents, each with HFA, OO, or TD, were selected from a larger study of OO such that groups did not differ on age or IQ. Participants from Study 1 were not included in Study 2, creating a distinct sample. Demographic information is shown in Table 4.

For inclusion, all participants were required to have VIQ, NVIQ, and FSIQ scores above 77 on the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). To be included in the HFA group, participants were required to meet criteria for diagnosis of ASD on the basis of the ADOS, parent report, and clinical judgment by an expert clinician. Participants in the OO group were required to have a documented diagnosis of ASD from a specialist in the field prior to age 5 years and to no longer meet diagnostic criteria for ASD on the basis of both the ADOS and clinical judgment. In addition, they had to be in regular-education classrooms with no special education services for ASD-related symptoms. Participants in the OO and TD groups were required to have average scores (>77) on the Communication and Social domains of the Vineland Adaptive Behavior Scales, Second Edition (Sparrow, Cicchetti, & Balla, 2005). Participants were excluded if they exhibited major psychopathology or hearing and vision deficits that would affect their ability to participate. Exclusion criteria also included seizures, fragile X disorder, and head trauma with a loss of consciousness. Fein et al. (2013) provide a description of the larger study, including the recruitment and screening procedures and inclusion criteria.

Table 4. Demographic information for participants with high-functioning autism (HFA), optimal outcomes (OO), and typical development (TD): Study 2.

Characteristic	HFA	OO	TD	*F*	*p*	Post hoc
Gender (M;F)	13;1	12;3	14;1			
Age (years)	13.0 (1.6), 10.5 to 15.7	13.0 (2.15), 9.3 to 17.1	12.9 (1.5), 9.9 to 15.6	0.1	.99	
WASI NVIQ	106.1 (13.1), 78 to 120	110.1 (15.2), 87 to 131	117.8 (12.8), 89 to 139	2.73	.08	
WASI VIQ	103.9 (13.9), 81 to 133	113.0 (18.1), 80 to 137	111.1 (13.6), 93 to 136	1.42	.25	
WASI FSIQ	105.5 (11.0), 80 to 117	112.8 (15.8), 82 to 134	115.5 (11.9), 101 to 142	2.22	.12	
CELF-4	102.1 (12.2), 88 to 124	111.0 (13.1), 79 to 126	119.6 (8.6), 106, to 132	9.14	.001	HFA < TD, OO
PPVT-3	106 (14.2), 86 to 130	114.5 (14.5), 86 to 129	122.2 (11.2), 104 to 143	4.09	.02	HFA < TD
ADOS Communication	3.3 (1.5), 2 to 7	0.47 (0.64), 0 to 2	0.2 (0.6), 0 to 2	38.93	.0001	HFA < TD
ADOS Social	6.8 (2.5), 4 to 13	0.80 (1.1), 0 to 3	0.4 (0.8), 0 to 2	71.05	.0001	HFA < TD, OO

Note. Data are shown as *M* (*SD*), range. WASI = Wechsler Abbreviated Scale of Intelligence; NVIQ = nonverbal IQ; VIQ = verbal IQ; FSIQ = full-scale IQ; CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition; PPVT-3 = Peabody Picture Vocabulary Test–III; ADOS = Autism Diagnostic Observation Schedule.

Prior to testing, written consent or assent was obtained from both parents and participants. This research was approved by all relevant institutional review boards.

Measures

As in Study 1, the ADOS was used as the gold standard assessment for verifying and ruling out ASD diagnosis. The Social Communication Questionnaire was also used as the parent-report measure of ASD symptomatology, and the PPVT-3 was used as a measure of lexical knowledge. Additional measures included the Wechsler Abbreviated Scale of Intelligence—an assessment of cognitive abilities used to obtain a measure of FSIQ, VIQ, and NVIQ—and the core subtests of the Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel, Wiig, & Secord, 2003)—an assessment of receptive and expressive language skills. These subtests were administered to evaluate receptive and expressive language functioning of participants, which is used to diagnose language and communication disorders.

Procedure

Participants were tested in a quiet room at their home, the University of Connecticut, or the Institute of Living of Hartford Hospital. The measures included in this study were also collected as part of a larger study, completed in two or three sessions.

Participants completed the same experimental task as in Study 1, the ADOS Monkey Cartoon task, with one methodological difference: In Study 1, the picture cards were removed from view before the participants told the story, whereas in Study 2 they remained in view during the narration. All narratives were video-recorded for later transcription and analysis. Identical transcription and coding procedures were used as in Study 1.

Reliability

The first author coded all 45 narratives for both completeness and story grammar. A trained research assistant coded 33.3% of the transcripts ($n = 15$) to obtain reliability data. Both coders were unaware of diagnoses. Analyses indicated excellent interrater reliability across all categories (story completeness: ICC = .81; story grammar: ICC = .93; setting: ICC = .96; internal response: ICC = .86; and added/invented detail: ICC = .95).

Story-Quality Ratings

Forty-seven University of Connecticut undergraduate students rated narrative quality in return for participation credit in a psychology class. Raters were older than 18 years of age and were native speakers of English. Eight raters were excluded due to poor performance on the narrative task.

The raters completed the same rating procedure as in Study 1, with one difference: In Study 1, raters evaluated 10 narratives, whereas in Study 2, raters evaluated 12. The order of narratives was randomized for each rater.

Results

Statistical analyses compared the performance of the HFA, OO, and TD groups on several measures of overall story goodness as well as on more specific components of narrative quality. The procedures for the data analysis in Study 2 were identical to those in Study 1. One participant with HFA was removed from further analyses due to exceptionally poor experimental-task performance, because his story-completeness and story-grammar scores were both more than 2 *SD*s below the mean. No other data were excluded from further analyses. Table 5 presents coding-category data.

Narrative Analysis

Story composite score. A univariate analysis of variance was conducted for the story composite score. Scores differed significantly by group, $F(2, 41) = 4.24$, $p = .02$, $\eta_p^2 = .17$. Post hoc tests indicated that the contrast between the HFA and TD groups was significant, $F(1, 27) = 11.75$, $p = .002$, $\eta_p^2 = .30$, with the TD group having higher scores than the HFA group. The OO and HFA groups' scores did not differ, $F(1, 27) = 2.19$, $p = .15$, $\eta_p^2 = .08$, nor did the OO and TD groups', $F(1, 28) = 1.57$, $p = .22$, $\eta_p^2 = .05$.

Table 5. Coding results as a function of group: Study 2.

Measure	HFA	OO	TD	*F*	*p*	η_p^2	Post hoc
Story composite[a]	–0.30 (0.53), –1.56 to 0.68	0.02 (0.65), –1.26 to 1.11	0.26 (0.35), –0.32 to 0.84	4.24	.02	.17	HFA < TD
Story completeness[b]	3.9 (1.2), 1 to 5	4.4 (1.7), 1 to 6	4.8 (1.2), 2 to 6	1.75	.19	.08	
Story grammar[c]	.62 (.14), .29 to .83	.60 (.22), .17 to 1.0	.64 (.13), .40 to .88	0.18	.84	.01	
Coding categories[d]							
Setting	0.43 (0.76), 0 to 2	0.67 (0.62), 0 to 2	1.27 (0.96), 0 to 4	4.36	.02	.18	HFA < TD
Internal response	0.79 (0.89), 0 to 3	1.00 (0.65), 0 to 2	1.33 (1.04), 0 to 3	1.44	.25	.07	
Added/invented detail	1.0 (1.30), 0 to 4	1.87 (1.88), 0 to 6	0.87 (1.25), 0 to 4	1.93	.16	.09	
Total T-units	8.4 (1.8), 6 to 12	10.9 (4.7), 4 to 20	10.0 (2.3), 6 to 14	2.14	.13	.10	

Note. Data are shown as *M* (*SD*), range. HFA = high-functioning autism; OO = optimal outcomes; TD = typical development.

[a]Average z-scores. [b]Average score out of 6 total. [c]Average proportion (i.e., sum of T-units that fell into the categories of initiating event, attempt, and direct consequences divided by the total number of T-units in the narrative). [d]Average frequency.

Overall story goodness is depicted in Figure 2, which plots story completeness against story grammar.

Story completeness. The completeness scores represent the number of key story events expressed in the participant's narrative. These scores showed no group differences, $F(2, 41) = 1.75$, $p = .19$, $\eta_p^2 = .08$.

Story grammar. The story grammar scores represent the proportion of T-units within story-grammar episode structure (i.e., initiating events, attempts, and direct consequences). The story grammar scores did not differ by group, $F(2, 41) = 0.18$, $p = .84$, $\eta_p^2 = .01$.

Additional coding categories. To follow up on the significant group difference for the story composite scores, analyses probed setting and internal response separately, as well as added/invented detail and total T-units. There was a group difference in setting, $F(2, 41) = 4.36$, $p = .02$, $\eta_p^2 = .18$. Post hoc tests indicated that the difference was driven by group differences for the HFA and TD groups, $F(1, 27) = 6.75$, $p = .02$, $\eta_p^2 = .20$. There were no main effects of Group for other coding categories; results are shown in Table 6.

Story-Quality Ratings

Undergraduate students rated 12 narratives each. Interrater reliability was calculated using one-way random average ICC. Results indicated that average interrater reliability ranged from good to excellent (goodness: ICC = .87; cohesiveness: ICC = .81; accuracy: ICC = .86; oddness: ICC = .66).

A repeated measures analysis of variance on story-quality ratings revealed a significant main effect of Group, $F(1, 41) = 1860.18$, $p < .001$, $\eta_p^2 = .98$, no main effect of Dimension Rating, $F(3, 123) = 2.33$, $p = .08$, $\eta_p^2 = .06$, and a significant Group × Dimension Rating interaction, $F(6, 123) = 4.60$, $p < .001$, $\eta_p^2 = .18$. Follow-up analyses were conducted for story goodness, cohesiveness, accuracy, and oddness; results are shown in Table 6. Ratings data indicated that the HFA group received significantly lower ratings than the TD group for story goodness, cohesiveness, and accuracy and significantly higher ratings for oddness. Although the OO group did not differ from the HFA or TD groups on accuracy or oddness ratings, they did receive significantly lower ratings than the TD group on story goodness and cohesiveness. Correlations between goodness ratings, story composite scores, and age were examined, but none were significant.

Discussion

Study 2 assessed pragmatic language skills during narrative production in adolescents with HFA, OO, and TD using the story-goodness coding system and ratings of narrative quality. This study extended findings of Study 1 to a sample of individuals with OO.

In Study 2, there were no group differences between the HFA, OO, and TD groups on the measures of story grammar, mental state language, added/invented details, or narrative length. This effect, replicated from Study 1

Figure 2. Study 2 story goodness. Story goodness was plotted as a function of story grammar and story completeness. Again, lines dividing the graphs into quadrants are set at 1 *SD* below the group mean for typical development. The narrative performance of the groups with high-functioning autism spectrum disorder and optimal outcomes is more varied than that of the group with typical development.

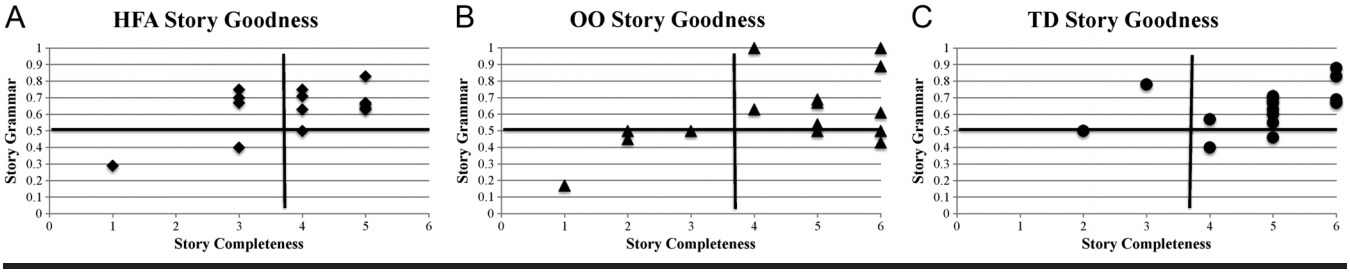

Table 6. Narrative-quality ratings: Study 2.

Dimension	HFA	OO	TD	*F*	*p*	η_p^2	Post hoc
Goodness	2.15 (0.58), 1.2 to 3.0	2.61 (0.89), 1.5 to 4.1	3.27 (0.74), 2.0 to 4.5	8.10	.001	.28	HFA < TD, $p < .001$; OO < TD, $p = .04$
Cohesiveness	2.71 (0.73), 1.2 to 4.0	2.89 (0.89), 1.7 to 4.3	3.53 (0.73), 2.3 to 4.6	4.47	.02	.18	HFA < TD, $p = .01$; OO < TD, $p = .04$
Accuracy	2.53 (0.68), 1.4 to 3.8	2.83 (0.96), 1.5 to 4.3	3.47 (0.89), 1.7 to 4.5	4.71	.01	.19	HFA < TD, $p = .003$
Oddness	3.05 (0.55), 2 to 4	2.85 (0.66), 1.9 to 3.9	2.49 (0.59), 1.7 to 3.8	3.33	.05	.14	HFA > TD, $p = 0.01$

Note. Data are shown as *M* (*SD*), range. Scores are represented as average ratings on a 5-point scale. HFA = high-functioning autism; OO = optimal outcomes; TD = typical development.

with a new sample, indicates that story structure (e.g., story grammar) is a relative strength in HFA, consistent with findings by Young et al. (2005). This is also consistent with other previous studies of autism narratives that have failed to find differences in narrative macrostructure when wordless picture books were used as stimuli (Diehl et al., 2006; Losh & Capps, 2003). Contrary to the findings in Study 1, the HFA group did not show a trend toward less complete narratives than the TD group.

The HFA group again received significantly lower scores than the TD group on a broad measure of narrative quality, the story composite score, with the OO group not differing from either other group. The HFA, OO, and TD groups performed similarly on most specific measures of narrative quality, with only one difference between the HFA and TD groups: The HFA group provided significantly less setting information. As in Study 1, this finding suggests that the combination of impairments in several narrative skills leads to diminished narrative quality in HFA.

The HFA and TD groups differed in all four rating domains: goodness, cohesiveness, accuracy, and oddness. These findings are particularly robust in that group differences were detectable by naïve raters from reading transcribed narrations, without the influence of prosodic or other personal characteristics that might affect perceptions of narrative quality. Furthermore, the OO and TD groups differed on ratings of goodness and cohesiveness. Although individuals with OO are virtually indistinguishable from their peers with TD in many ways, they show continued subtle difficulty in pragmatic language (Kelley et al., 2006; Suh et al., 2014). This difficulty was apparent to naïve raters but was not detected in story coding. In addition, plots depicting story goodness as the combination of story grammar and story completeness highlighted the individual variability in narrative performance in the HFA and OO groups. Although pragmatic language in HFA is generally recognized as an important locus of intervention, adolescents with OO may also benefit from targeted intervention in narrative language.

General Discussion

The aim of this research was to use a story-coding system to characterize pragmatic language strengths and weaknesses in individuals with HFA and OO from ASD. Across both studies, with short narratives, story composite scores were significantly lower for participants in the HFA group as compared with participants in the TD group, whereas participants in the OO group did not differ from either other group; in contrast, story-grammar or narrative-organization abilities were similar in adolescents with HFA, OO, and TD. The findings also suggest that a broad measure of narrative quality (i.e., story composite score) best captures the deficits in narrative ability observed in HFA, because it is not a single element of narrative production in isolation that drives the differences in narrative quality but rather a constellation of slight impairments across several components of narrative production. These differences in narrative-quality scores were supported by the similar narrative-quality judgments made by naïve raters; participants with HFA received lower ratings on average than participants with TD. It is interesting to note that although participants with OO did not significantly differ from participants with HFA or TD in the broad story composite measure, they did receive significantly lower ratings of goodness and cohesiveness than the TD group, indicating the presence of persisting subtle pragmatic differences in the OO group.

In Study 1, the participants with HFA had less complete narratives as compared with their peers with TD, whereas this difference was not apparent in Study 2. In Study 1, where the picture stimuli were removed during narrative production, the narratives of the HFA group tended to be similar in length to those of the TD group yet less complete, which suggests that HFA narratives are less "on target," with a trend to produce more nonessential detail (consistent with findings from a study of picture descriptions; see Fitch, Fein, & Eigsti, 2015). In Study 2, when the picture stimuli cards were present, the completeness deficit in the HFA group disappeared. This indicates that it may be the working memory demands of Study 1 that contributed to the decreased recall of core plot components. Future research directly comparing the same group of participants on a narrative task with and without the presence of picture stimuli would help to tease apart the role of working memory in story construction.

These findings inform the assessment of pragmatic language abilities in individuals with HFA. Many standardized measures do not capture subtle group differences;

narrative elicitation allows for direct examiner observation of pragmatic abilities and serves as an ecologically valid metric of pragmatic abilities that is salient to naïve raters.

Limitations

There were several limitations relevant to both Study 1 and Study 2. The narrative samples were relatively brief, because the narrative stimuli were only six cards. These methods differ from those of many other studies in which participants are asked to narrate a longer picture book; these shorter narratives may have less power to detect group differences. Small sample sizes in both studies also reduced our power to detect group differences, though the consistency of results across studies suggests that effects are robust. Perhaps the most critical limitation involves structural language knowledge. In both Study 1 and Study 2, the HFA group had significantly lower scores (though within the average range) on standardized measures of receptive vocabulary (i.e., PPVT-3 in both studies) and other language fundamentals (i.e., Clinical Evaluation of Language Fundamentals–Fourth Edition in Study 2). Although groups did not differ in VIQ in either study, this group difference in more sensitive measures may have contributed to some of the narrative findings, and is a limitation in the current design. With PPVT-3 score as a covariate, the results were largely similar; we were reluctant to include different covariates across the two studies, which would impede comparison across studies. In fact, we would suggest that the group differences in narrative ability may actually reflect limitations in both pragmatic and structural language abilities in ASD; that said, group differences do somewhat limit interpretation of the findings.

These studies indicate that story grammar is an area of relative strength in HFA, whereas other aspects of narrative quality, such as completeness, may be constrained by working memory demands. Ratings of narrative quality in HFA and TD revealed group differences, indicating that pragmatic language deficits displayed in narrative production are likely salient in everyday communication. Furthermore, the detection of differences between the OO and TD groups on ratings of narrative quality suggest that subtle pragmatic deficits may remain even though these adolescents with OO are performing as well as their peers in many other language, social, and adaptive domains.

Acknowledgments

This research was supported by the National Institute of Mental Health Grant R01MH076189 (Deborah Fein), University of Connecticut Research Foundation Grant #458938 (Inge-Marie Eigsti), and National Institute of Neurological Disorders and Stroke Grant T32NS007413 (Ashley de Marchena).

We would like to thank the participating children and families who made this research possible, as well as contributing members of the Developmental Cognitive Neuroscience and Optimal Outcome Labs for their assistance with data collection and coding, including Emily Thompson, Catherine Piotrowski, and Christopher Andrade.

References

Abbeduto, L., Benson, G., Short, K., & Dolish, J. (1995). Effects of sampling context on the expressive language of children and adolescents with mental retardation. *Mental Retardation, 33,* 279–288.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

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

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4,* 113–125.

Beaumont, R., & Newcombe, P. (2006). Theory of mind and central coherence in adults with high-functioning autism or Asperger syndrome. *Autism, 10,* 365–382.

Bishop, D. (2006). *Children's Communication Checklist-2*. San Antonio, TX: The Psychological Corporation.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching & Therapy, 18,* 1–21. doi:10.1191/0265659002ct224oa

Boudreau, D. (2008). Narrative abilities: Advances in research and implications for clinical practice. *Topics in Language Disorders, 28,* 99–114.

Capps, L., Losh, M., & Thurber, C. (2000). "The frog ate the bug and made his mouth sad": Narrative competence in children with autism. *Journal of Abnormal Child Psychology, 28,* 193–204.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290.

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

Dale, P. S. (1996). Parent report assessment of language and communication. In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language* (Vol. 6; pp. 161–182). Baltimore, MD: Brookes.

Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society, 15,* 331–343.

de Marchena, A., & Eigsti, I.-M. (2016). The art of common ground: Emergence of a complex pragmatic language skill in adolescents with autism spectrum disorders. *Journal of Child Language, 43,* 43–80.

Diehl, J. J., Bennetto, L., & Young, E. C. (2006). Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *Journal of Abnormal Child Psychology, 34,* 83–98. doi:10.1007/s10802-005-9003-x

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–III*. Circle Pines, MN: AGS.

Eigsti, I.-M., & Bennetto, L. (2009). Grammaticality judgments in autism: Deviance or delay. *Journal of Child Language, 36,* 999–1021.

Eigsti, I.-M., Bennetto, L., & Dadlani, M. B. (2007). Beyond pragmatics: Morphosyntactic development in autism. *Journal of Autism and Developmental Disorders, 37,* 1007–1023.

Eigsti, I.-M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders, 5,* 681–691. doi:10.1016/j.rasd.2010.09.001

Fein, D., Barton, M., Eigsti, I.-M., Kelley, E., Naigles, L., Schultz, R. T., . . . Tyson, K. (2013). Optimal outcome in individuals with a history of autism. *The Journal of Child Psychology and Psychiatry, 54,* 195–205.

Fitch, A., Fein, D. A., & Eigsti, I.-M. (2015). Detail and gestalt focus in individuals with optimal outcomes from autism spectrum disorders. *Journal of Autism and Developmental Disorders, 45,* 1887–1896. doi:10.1007/s10803-014-2347-8

Fox, B. J., & Wright, M. (1997). Connecting school and home literacy experiences through cross-age reading. *The Reading Teacher, 50,* 396–403.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8,* 8–11.

Hartley, L. L., & Jensen, P. J. (1991). Narrative and procedural discourse after closed head injury. *Brain Injury, 5,* 267–285.

Hauerwas, L. B., & Stone, C. A. (2000). Are parents of school-age children with specific language impairments accurate estimators of their child's language skills? *Child Language Teaching & Therapy, 16,* 73–86.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (NCTE Research Report No. 3). Champaign, IL: National Council of Teachers of English.

Kelley, E., Paul, J. J., Fein, D., & Naigles, L. R. (2006). Residual language deficits in optimal outcome children with a history of autism. *Journal of Autism and Developmental Disorders, 36,* 807–828. doi:10.1007/s10803-006-0111-4

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes, 16,* 287–308.

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

Landa, R., Martin, M., Minshew, N., & Goldstein, G. (1995, April). *Discourse and abstract language ability in non-retarded individuals with autism*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Indianapolis, IN.

Lê, K., Coelho, C., Mozeiko, J., & Grafman, J. (2011). Measuring goodness of story narratives. *Journal of Speech, Language, and Hearing Research, 54,* 118–126.

Liles, B. Z., Coelho, C. A., Duffy, R. J., & Zalagens, M. R. (1989). Effects of elicitation procedures on the narratives of normal and closed head-injured adults. *Journal of Speech and Hearing Disorders, 54,* 356–366.

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research, 38,* 415–425.

Lord, C., & Paul, R. (1997). Language and communication in autism. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (2nd ed.; pp. 195–225). New York, NY: Wiley.

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

Losh, M., & Capps, L. (2003). Narrative ability in high-functioning children with autism or Asperger's syndrome. *Journal of Autism and Developmental Disorders, 33,* 239–251.

Loveland, K. A., McEvoy, R. E., Tunali, B., & Kelley, M. L. (1990). Narrative story telling in autism and Down's syndrome. *British Journal of Developmental Psychology, 8,* 9–23.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk—Vol. II: The database* (3rd ed.). Mahwah, NJ: Erlbaum.

Merritt, D. D., & Liles, B. Z. (1987). Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension. *Journal of Speech and Hearing Research, 30,* 539–552.

Mozeiko, J., Lê, K., Coelho, C., Krueger, F., & Grafman, J. (2011). The relationship of story grammar and executive function following TBI. *Aphasiology, 25,* 826–835.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders, 38,* 287–313.

Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language, 41,* 485–510.

Phelps-Terasaki, D., & Phelps-Gunn, T. (1992). *Test of Pragmatic Language*. Austin, TX: Pro-Ed.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition*. Austin, TX: Pro-Ed.

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

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

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales, Second Edition (Vineland-II)*. Circle Pines, MN: AGS.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Norwood, NJ: Ablex.

Suh, J., Eigsti, I.-M., Naigles, L., Barton, M., Kelley, E., & Fein, D. (2014). Narrative performance of optimal outcome children and adolescents with a history of an autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders, 44,* 1681–1694.

Tager-Flusberg, H., Joseph, R., & Folstein, S. (2001). Current directions in research on autism. *Mental Retardation and Developmental Disabilities Research Reviews, 7,* 21–29.

Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics, 16,* 241–256. doi:10.1017/S0142716400007281

Thorndyke, P. W., & Yekovich, F. R. (1980). A critique of schema-based theories of human story memory. *Poetics, 9,* 23–49.

Tyson, K., Kelley, E., Fein, D., Orinstein, A., Troyb, E., Barton, M., . . . Rosenthal, M. (2014). Language and verbal memory in individuals with a history of autism spectrum disorders who have achieved optimal outcomes. *Journal of Autism and Developmental Disorders, 44,* 648–663. doi:10.1007/s10803-013-1921-9

Volden, J., & Phillips, L. (2010). Measuring pragmatic language in speakers with autism spectrum disorders: Comparing the Children's Communication Checklist-2 and the Test of Pragmatic Language. *American Journal of Speech-Language Pathology, 19,* 204–212.

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

Young, E. C., Diehl, J. J., Morris, D., Hyman, S. L., & 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.

Appendix

Sample Coded Transcript

Participant: typically developing, age 13
Story-grammar score: .88
Story-completeness score: 6/6
Narrative length: 8 T-units

1. There's a monkey who was collecting coconuts. *(Attempt)*
2. and he dropped one on the ground. *(Initiating Event)*
3. and the monkey below him took it and ran away with it. *(Attempt)*
4. the monkey who was getting coconuts didn't know what had happened to the coconut. *(Internal Response)*
5. so he dropped another and watched it. *(Initiating Event)*
6. and the monkey that was below him before came back and took it. *(Attempt)*
7. and when the monkey below him was about to run away, the monkey in the tree threw a coconut. *(Direct Consequence)*
8. and it hit the monkey below him in the head. *(Direct Consequence)*

See Merritt and Liles (1987) for the full coding criteria for the following categories: initiating event, attempt, direct consequence, setting, and internal response. Added/invented detail was coded for statements that describe story events or excessive details not present in the original story (the picture cards), such as additional plans, conflicts, thoughts, wishes, and so on.