University of Connecticut
OpenCommons@UConn

Master's Theses

University of Connecticut Graduate School

6-6-2014

"Story Goodness" in Adolescents with Autism Spectrum Disorders

Allison R. Canfield
University of Connecticut, allison.canfield@uconn.edu

"Story Goodness" in Adolescents with Autism Spectrum Disorders

Allison R. Canfield

B.S., University of Rochester, 2010

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

at the

University of Connecticut

2014

APPROVAL PAGE

Master of Arts Thesis

"Story Goodness" in Adolescents with Autism Spectrum Disorders

Presented by

Allison R. Canfield, B.S.

Major Advisor _____
Inge-Marie Eigsti, Ph.D.

Associate Advisor _____
Deborah Fein, Ph.D.

Associate Advisor _____
Carl Coelho, Ph.D.

University of Connecticut

2014

Acknowledgements

Table of Contents

Abstract

Individuals with autism spectrum disorders (ASD) are characterized by pragmatic language deficits, as difficulties with social communication are a universal deficit across the wide presentation the autism spectrum. The current studies examined narrative production as a measure of pragmatic language in adolescents with high-functioning autism (HFA), "optimal outcomes" (OO; i.e. individuals who have lost their previous diagnosis of ASD), and typical development (TD) using the informative "story goodness" coding system and real-world narrative quality ratings. In Study 1, adolescents with HFA and TD (mean age of 15) were compared on these measures of narrative quality. While the groups were similar in their narrative organization as measured by story grammar, the lower completeness scores of the HFA narratives trended towards significance. Raters naïve to diagnoses and study hypotheses rated the HFA group as having significantly lower scores of Story Cohesiveness and trended towards significantly lower scores of Story Goodness. In Study 2, the narrative quality of adolescents with HFA, OO, and TD (mean age of 13) was compared. While group differences were not seen on measures of story completeness or story grammar, naïve raters did detect group differences between HFA and TD in all four rating domains: Story Goodness, Story Cohesiveness, Story Accuracy, and Story Oddness, and between OO and TD in Story Goodness and Story Cohesiveness. In both Study 1 and Study 2, group differences emerged between HFA and TD in a broad measure of narrative quality, with the HFA group receiving significantly lower scores on this metric. These findings highlight the presence of pragmatic deficits in adolescents with HFA that are salient in the "real world" despite their comparable performance to their TD peers in other language domains. Furthermore, the findings support prior research that indicates while individuals with OO are similar to their TD peers across several domains (i.e. cognitive,

achievement, and adaptive functioning), subtle pragmatic deficits remain in these individuals as evident in their narrative performance.

"Story Goodness" in Adolescents with Autism Spectrum Disorders

Autism spectrum disorders (ASD) are a heterogeneous group of neurodevelopmental disorders that are characterized by impairments in social interaction and communication, including spoken language, as well as restricted and repetitive behaviors and interests (American Psychiatric Association, 2000). Pragmatic language, or socially appropriate communication, lies at the intersection of the social and communication deficits in ASD, and as such, is the most consistent domain of deficit. Pragmatic language deficits impede social interactions and the development of peer relationships. Furthermore, some pragmatic deficits remain in children who have achieved "optimal outcomes" and no longer meet diagnostic criteria for ASD (Kelley, Paul, Fein, & Naigles, 2006). Social communication deficits in ASD warrant further investigation to illuminate underlying mechanisms and specific areas of deficit within the broad domain of pragmatic language; this is a primary focus of the current paper.

Pragmatic language includes aspects of verbal communication, such as appropriate word choice and content, turn-taking in conversations, and prosody, which is the rhythm, stress, and intonation of speech. It also includes nonverbal behaviors that are socially relevant to communication, such as eye contact, gestures, and facial expressions (Eigsti, de Marchena, Schuh, & Kelley, 2011). As ASD is characterized by impairments in the recognition of social cues, as well as difficulties with language production and comprehension, individuals with ASD by definition exhibit impairments in pragmatic language (Bishop, 1989).

The presentation of language abilities in ASD can vary greatly, with some individuals remaining nonverbal (Lord & Paul, 1997) and others acquiring language abilities that are within normal limits (Landa, 2000). In children with ASD who do acquire language, certain language

skills, such as vocabulary and grammar, may develop later in ASD than in those with typical

development (TD) (Eigsti & Bennetto, 2009; Eigsti, Bennetto & Dadlani, 2007), but can often

develop to be in the average or above average range according to standardized tests (Kjelgaard &

Tager-Flusberg, 2010; Tager-Flusberg, Joseph, & Folstein, 2001). Despite the later acquisition of

average verbal abilities, the social communication impairments do not show the same pattern

(Landa, 2000; Lord & Paul, 1997; Tager-Flusberg, 2004); therefore, pragmatic abilities are

considered to be the most universal deficit across the autism spectrum.

While there are many standardized assessments of language fundamentals, such as

vocabulary, semantics, and syntax, assessments of pragmatic language deficits are limited

(Young, Diehl, Morris, Hyman, & Bennetto, 2005). Such tests have proven difficult to design as

much of social communication is "context-dependent" and requires flexibility, both of which are

difficult to evaluate in standardized assessments (Volden & Philips, 2010). Indeed, the highly

structured format of standardized tests is likely why these tests are often found to be a strength in

individuals with ASD (Kelley et al., 2006). The best way to capture and characterize pragmatic

language skills across clinical populations continues to be an ongoing debate. For instance,

previous research has indicated that standardized measures of pragmatic language, such as the

Test of Pragmatic Language (TOPL; Phelps-Terasaki & Phelps-Gunn, 1992), can adequately

capture the poor performance of individuals with ASD (Young et al., 2005), while Volden and

Philips (2010) reported that a parent-report measure of social communication skills, the

Children's Communication Checklist-2 (CCC-2, Bishop, 2006) better differentiated ASD and

TD groups than the TOPL. While the CCC-2 may be a more useful assessment tool for

pragmatic abilities in ASD than other established measures, research has indicated that parent-

report of language abilities may not be as accurate in school-aged children and adolescents as it

is in young children, when their language abilities are still emerging (Dale, 1996; Boyton-Hauerwas & Stone, 2000) This highlights the need for an assessment tool that allows for direct examiner observation of pragmatic language abilities in order to best evaluate pragmatic strengths and weaknesses in older children and adolescents.

Narratives as a Measure of Pragmatic Language

Due to limitations in the standardized assessments of pragmatic abilities, narrative production is commonly used to characterize both language and social abilities across fields. To elicit a narrative sample, participants produce a fictional or non-fictional story, either from memory or by narrating a story (often in response to picture stimuli), and the spoken stories are subsequently transcribed and coded. The production of a spoken narrative is a complex task, tapping semantic, grammar, and pragmatic abilities, as well as executive functioning, as coherent narrative production requires planning and organization (Diehl, Bennetto, & Young, 2006). In many ways, narrative production is a more complex and demanding task than engaging in conversations (Abbeduta, Benson, Short & Dolish, 1995; Botting, 2002), as conversational partners provide priming and structure for the speaker to follow (Garrod & Pickering, 2004). Syntactic or grammatical complexity, diversity of words used, and overall narrative length increase as a child's language skills develop with age (Botting, 2002). Both the comprehension and production of narratives contribute to academic success and effective social interaction in children with TD (Boudreau, 2008; Fox & Wright, 1997; Houston, 1997). Therefore, narrative abilities seem to capture real-world social skills and abilities that are central to pragmatic language.

Narratives in ASD. Narrative elicitation is a useful tool with which to study pragmatic language in ASD. Many of our daily interactions involve the narrative form and as such,

narrations are an ecologically valid language sample. Narratives have been used to examine narrative length, structure, and cohesion in individuals with ASD. In studies of ASD and TD groups that were well-matched on language and cognitive abilities, results suggest minimal quantitative differences on broad narrative measures, such as length and syntactic complexity (Diehl et al., 2006; Tager-Flusberg & Sullivan, 1995); yet, children with ASD produce narratives with less organization and cohesiveness. That is, their narratives tend lack coherent global organization of the main ideas and central events (Cohen & Leslie, 1986; Loveland, McEvoy, & Tunali, 1990; Suh et al., 2014), and their narratives are significantly less cohesive (i.e. fewer causal connections between story events) and not as well organized around the overall "gist" story components (Diehl, Bennetto, & Young, 2006).

In addition to story organization, an important metric of narrative quality is its completeness. In a picture book narration study, Suh and colleagues (2014) reported that children with high-functioning autism (HFA; individuals with ASD and IQs above 80) provided less complete narratives than a TD group matched on age that did not significantly differ in gender or verbal IQ. This completeness deficit appears to persist into late adolescence and adulthood; Landa, Martin, Minshew, and Goldstein (1995) found that adults and adolescents with high functioning autism produced narratives that were similar in length (e.g., number of independent clauses) to the TD comparison group, but were less complete (e.g., significantly more incomplete story episodes) in their retelling. Additionally, a recent study by Norbury, Gemmell, and Paul (2013) reported that ASD children in their study produced relatively simplistic narratives characterized by more pragmatic errors, a less cohesive narrative macrostructure, and the omission of significant story elements.

Narrative studies have also been used to examine *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). These abilities may relate to theory of mind, a deficit characteristic of ASD in which affected individuals have difficulty understanding the thoughts and feelings of others. While 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, 2009; Tager-Flusberg & Sullivan, 1995).

Optimal Outcomes from ASD. Historically, ASD has been seen as a lifelong disorder; however, a growing literature has explored the phenomenon of "optimal outcome" (OO) where individuals with a previous diagnosis of ASD no longer meet the diagnostic criteria for the disorder. Fein and colleagues (2013) found that children with OO were indistinguishable from a well-matched group of typically developing peers on standardized measures of language, socialization, and adaptive skills. Furthermore, children with OO have been found to perform similarly to matched groups of children with TD on several standardized assessments of language abilities (Tyson et al., 2013) and in areas of academic achievement (Troyb et al., 2013) where their peers with HFA continue to show deficits. However, Kelley and colleagues (2006) reported that individuals with OO at age eight displayed subtle deficits on some pragmatic language measures (i.e. theory of mind, narrative performance and use of mental state verbs) despite their otherwise normative language performance.

Few studies have examined the narrative performance of individuals with OO. Kelley and colleagues (2006) reported the OO group with a mean age of eight produced significantly fewer causal connections, less clear referents, and less discussion of the goals and motivations of the characters in picture book narrations than typically developing controls. An analysis of narratives

by OO with a mean age of 12 revealed deficits in two specific domains: speech dysfluencies and idiosyncratic language (Suh et al., 2014). The OO participants were 9 times more likely than the TD group to use idiosyncratic language, such as highly specific or unusual references to topics outside the scope of the story, in their story telling. This indicates that 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 in their verbal skills.

Narratives in TBI. Narratives have been used to examine how language abilities and organization may break down following a traumatic brain injury (TBI) and to highlight both the gross and more subtle communication deficits that may occur (Hartley & Jensen, 1991; Liles, Coelho, Duffy, & Zalagens, 1989; McDonald, 1993; Mentis & Prutting, 1987; Mozeiko, Le, Coelho, Krueger, & Grafman, 2011). Deficits include producing less complete, organized and cohesive narratives (Coelho, Liles, & Duffy, 1995). As in the autism literature, a wide variety of narrative elicitation procedures and analyses has been used, making it difficult to draw comparisons across studies and clinical populations (Coelho, 2002; Diehl et al., 2006).

Prior studies have utilized "story grammar", or the organization of story components into logical relationships between characters and other story events, as a means to assess narrative quality or richness (Stein & Glenn, 1979; Thorndyke & Yekovich, 1980). Story grammar quantifies the plot components that are structured around prototypical story structures, such as a character's goals; initiating actions towards these goals; attempts at solutions to problems; and direct consequences of attempts that result in the attainment or nonattainment of the goal (Liles et al., 1989; Liles, Duffy, Merritt, & Purcell, 1995). Story grammar can highlight the cognitive, semantic and pragmatic impairments, including poorly organized narratives, that may follow TBI (Lê, Coelho, Mozeiko, & Grafman, 2011a).

Lê, Coelho, Mozeiko, and Grafman (2011a) proposed a novel metric of "story goodness" that integrates measures of narrative completeness and narrative organization (i.e. story grammar). Participants were asked to view a 16-frame wordless cartoon story and then to retell the story. Results showed that the narratives of adults with TBI were significantly less complete and organized than in the comparison group. This story goodness measure better characterized performance of the TBI group than looking at either story completeness or organization measures in isolation. While previous studies of ASD have included elements of story grammar (Young et al., 2005), none have yet used this informative "story goodness" measure. Narrative analysis using the story goodness coding system will allow for comparisons between ASD, TBI, and other clinical populations, and may highlight subtle pragmatic language deficits even in a high-functioning ASD sample with generally excellent language abilities.

Current Study

The present study evaluated pragmatic language abilities and narrative quality across two samples: 1) adolescents with HFA and TD and 2) adolescents with OO, HFA and TD, using the story goodness coding system (Lê, Coelho, Mozeiko, and Grafman, 2011a). Based on previous research, we predicted that, compared to a TD comparison group, stories by adolescents with HFA would be of similar length, but would be less complete. In contrast, we predicted that individuals with OO would not show this deficit. On a measure of narrative organization (i.e. story grammar), we predicted the performance in this domain to be similar across groups, similar to the findings of Young and colleagues (2005). To evaluate the ecological validity of the story goodness metric, we acquired real-world ratings of narrative quality from naïve raters. On these measures, we predicted that participants with HFA would receive significantly lower scores on ratings of story goodness, cohesiveness, and oddness, but not on story accuracy. Overall, we

predicted that the narratives from the HFA groups would yield lower scores of "story goodness" compared to their TD peers; and that narratives from the OO group would be similar in quality to those of the TD group.

Study 1

Methods

Participants

Twenty adolescents with HFA and 16 adolescents with TD participated in this study. Participants were matched on chronological age, gender, and full-scale IQ as measured by the Stanford-Binet Intelligence Scales, Fifth Edition (Roid, 2003). Participants were recruited via fliers in the community 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 within the sample. Participants were excluded from the TD group if they had a history of neurological problems or any first-degree relatives with an 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 participant from the TD group were excluded from further analyses. The final sample included 15 adolescents with HFA and 15 with TD. Demographic information and average scores are 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 (Lord, Rutter, DiLavore, & Risi, 2002), the Social Communication Questionnaire (Rutter, Bailey, & Lord,

2003), and DSM-IV criteria (American Psychiatric Association, 2000). Within the HFA group, nine individuals met the diagnostic criteria for Autistic Disorder, five for PDD-NOS, and one for Asperger's Disorder. All participants in the HFA group were above the ADOS cutoff score of 7 for a diagnosis of an autism spectrum disorder or above the ADOS cutoff score of 10 for a diagnosis of autism, with the exception of one participant whose ADOS score was 3. This participant had a high score (20) on the Social Communication Questionnaire, consistent with a diagnosis of ASD, and the study's clinicians determined that this participant met DSM-IV criteria for an ASD diagnosis.

Prior to testing, written consent or assent was obtained from parents and participants. This research was approved by the University of Connecticut Institutional Review Board.

Measures

Stanford-Binet Scales, Fifth Edition (Roid, 2003). The Stanford-Binet is a measure of intelligence and cognitive abilities. The participants were assessed using the matrices and vocabulary subtests to obtain a measure of their overall cognitive functioning (full scale IQ; FSIQ), as well as their nonverbal (NVIQ) and verbal (VIQ) cognitive functioning.

Peabody Picture Vocabulary Test, Third Edition (Dunn & Dunn, 1997). The PPVT is a standardized measure of receptive vocabulary, which provides an estimate of vocabulary acquisition.

Social Communication Questionnaire (Rutter et al., 2003). The SCQ is a standardized parent-report measure that identifies symptoms of autism spectrum disorders and is designed to be an autism-diagnostic screening tool. All parents completed the Lifetime version. Data from twenty-seven parents was included as three parents (two in the HFA group, one in the TD group) did not return the questionnaire.

Autism Diagnostic Observation Schedule (Lord et al., 2002). The ADOS is a semi-structured assessment designed to diagnose autism spectrum disorders. Depending on the participant's maturity level, Module 3 (HFA group, $n = 6$; TD group, $n = 5$) or Module 4 (HFA group $n = 9$; TD group, $n = 10$) was administered; scores are provided in Table 1. All participants completed the Monkey Cartoon task of the ADOS, designed to assess the individual's verbal and nonverbal communication and their ability to integrate gestures with spoken language and to produce a coherent narrative. This task, in which participants produced a narrative based on a standard stimulus, served as the experimental measure for Study 1.

Table 1. Demographic Information for Participants with HFA and TD

	HFA *Mean (SD)* *Range*	TD *Mean (SD)* *Range*	χ^2 or *F*	*p*	η_p^2
N (M:F)	15 (13:2)	15 (14:1)	0.37	0.54	
Chronological age (years)	15.0 (1.5)	15.0 (1.5)	0.01	0.93	< 0.01
	12.4-17.3	12.8-17.6			
Stanford-Binet					
Nonverbal	10 (1.6)	11 (2.0)	2.68	0.11	0.09
	7 - 12	6 – 13			
Verbal	11 (2.6)	10 (1.6)	1.63	0.21	0.06
	6 - 16	7 – 13			
Full-scale IQ	103 (9.9)	103 (8.9)	0.003	0.95	< 0.01
	85 - 118	82 – 115			
PPVT	110 (12.8)	117 (8.0)	3.37	0.08	0.11
	78 - 127	105 - 137			
ADOS					
Communication	2.7 (1.0)	0.3 (1.0)	68.60	< 0.001	0.71
	1 - 4	0 – 1			
Social Reciprocity	7.3 (1.9)	0.3 (0.6)	184.19	< 0.001	0.87
	2 - 10	0 – 2			
C + SR	9.9 (2.6)	0.6 (0.6)	185.66	< 0.001	0.87
	3 - 13	0 – 2			
SCQ (total score)	20.6 (6.8)	1.9 (2.9)	88.09	< 0.001	0.78
	10 - 29	0 – 9			

Note: HFA, high functioning autism; TD, typically developing; PPVT, Peabody Picture Vocabulary Test; ADOS, Autism Diagnostic Observation Schedule; C + SR, Communication + Social Reciprocity; SCQ, Social Communication Questionnaire

Procedure

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

Stimuli for the narrative task were six black and white drawings (7 by 8.5 inches) depicting a story about two monkeys. The examiner instructed 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. If a participant had his

hands in his pockets, he was instructed to take them out; no further instructions were given. The

narrations were recorded on digital video.

Speech transcription. The narratives were transcribed from the videorecordings using

CLAN (MacWhinney, 2000). All words and dysfluencies were transcribed, and speech was

broken into minimal terminal units, known as T-units. A T-unit is a main clause and any attached

or embedded dependent clauses and is roughly equivalent to a sentence in spoken language

(Hunt, 1964; Liles, 1985).

Coding. The completeness and organization of narratives was coded using a system

described in Lê, Coelho, Mozeiko, and Grafman (2011a). *Story completeness* was

operationalized as the number of core story events, out of a possible six, expressed in a narrative.

A story event was essential to the comprehension of the plot; the six events are listed in Table 2.

Therefore, the *story completeness score* refers to the number of core story events expressed with

the scores ranging from 0 to 6.

Table 2. Story event descriptions.

Card	Event
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 him, he 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 back 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.

Story grammar or story organization was coded using procedures described by Le and colleagues (2011a), adapted in turn from Merritt and Liles (1987). Each t-unit was evaluated as falling into three categories: 1) *initiating events*, which prompt a character to act; 2) *attempts*, describing actions related to the initiating event; and 3) *direct consequences*, the results of attempts or the attainment or nonattainment of a goal. The *story grammar score* was calculated as the proportion of T-units that fell into one of these three categories, divided by the total number of T-units in the narrative, to control for narrative length (Lê et al., 2011a). A sample transcript is provided in Appendix A.

In addition to the three story grammar categories, there were three additional possible categories – *Setting* and *Internal Response* adapted from Merritt and Liles (1987) and *Added/Invented Detail*. *Setting* captured relevant information about a character's personality, habits or location. *Internal Response* comprised mental state descriptions of the thoughts and feelings of characters, similar to previous studies of mental state word use in ASD (Capps et al., 2000). Finally, *Added/Invented Detail* comprised off-topic or fictitious details that were included by some participants.

To characterize the quality of each narrative, we calculated a *Story Composite Score,* consisting of summed scores from six coding categories: *Story Completeness, Initiating Events, Attempts, Direct Consequences, Setting, and Internal Response*, each of which provides a measure of specific narrative qualities. The *Added/Invented Detail* and *Total T-units* categories were not included, as extraneous detail and excessive length detract from narrative quality.

Reliability. The first author, who was naïve to diagnosis, coded each of the thirty narratives for story completeness and story grammar. A trained research assistant, also naïve to diagnosis, coded 36% of the transcripts ($n = 11$) for interrater reliability, which was assessed at

88% for story completeness and 83% for story grammar in an utterance-by-utterance

comparison.

Story Quality Ratings. Thirty-nine University of Connecticut undergraduate students

rated narrative quality in return for extra credit in a psychology class. Raters were all 18 years of

age or older and were native speakers of English. Raters first viewed the six monkey cartoon

picture stimuli and told the story, which was recorded, and then read 10 transcribed narratives

(presented in random order, with five HFA and five TD narratives for each rater). Six raters were

excluded because their own narrative suggested poor comprehension of the picture cards or the

story. One participant was excluded because of a technical error.

The raters were naïve to the study's hypotheses. After reading a written transcription of a

narrative, raters answered four questions about each narrative, utilizing a 1-5 Likert scale.

Dysfluencies (i.e. "*uh*", "*um*", "*err*") and false starts (i.e. "the monkey came *ba-ba*-back") were

removed from narratives to prevent them from influencing ratings. Rating questions, presented in

this order, were: (1) *How good of a story is this?* (e.g., Story Goodness); (2) *How well were you*

able to follow this story? (e.g., Story Cohesiveness)*; (3) How well does this story reflect the*

actual story in the cards? (e.g., Story Accuracy); and (4) *How odd/unusual did you find this*

story? (e.g., Story Oddness). Narratives, rating questions and responses were presented and

collected using SuperLab software.

Results

To examine the narrative quality of adolescents with HFA, analyses compared the

performance of the HFA group to the TD group on several broad measures of "story goodness"

as well as on more specific components of narrative quality. Dependent variables were examined

for deviations from the assumptions of normality and sphericity, which were found to be

normally distributed. Data were checked for outliers (individuals with scores of more than 2 SD from the group mean). No data were excluded from further analyses. Effect sizes were calculated as partial eta squared (η^2_p), which refers to the proportion of variance attributable to a given effect after partialling out non-error sources of variance (Cohen, 1988).

Multivariate analysis of variance (MANOVA) was used to evaluate the six coding variables (i.e., *Story Completeness, Story Grammar, Setting, Internal Response, Added/Invented Detail*, and *Number of T-units*). All four multivariate tests were significant, $F(1, 28) = 305.20$, *p* < 0.001; Pillai's Trace = 0.99, Wilks' Lambda = 0.012, Hotelling's Trace = 79.62. Follow-up univariate analyses of variance (ANOVA) were used to probe for group differences on individual story measures. A summary of coding category data and statistical analyses, is provided in Table 3.

Table 3. Study 1 results.

	HFA	TD			
	M(SD)	*M(SD)*	*F*	*p*	η_p^2
Story Composite Score	10.9 (3.5)	14.4 (2.8)	9.25	0.005	0.25
Story Completeness	4.4 (1.5)	5.2 (0.8)	3.54	0.07	0.11
Story Grammar	0.60 (0.23)	0.62 (0.20)	0.07	0.80	0.002
Coding categories					
Setting	0.40 (0.83)	1.07 (1.33)	2.70	0.11	0.09
Internal Response	0.93 (0.88)	1.47 (1.41)	1.55	0.22	0.05
Added/Invented Detail	2.13 (3.50)	1.40 (1.64)	0.54	0.47	0.02
Total T-units	9.73 (4.13)	11.8 (4.50)	1.72	0.20	0.06

Narrative Analysis

Story Composite Score. To capture the combined contribution of each story component relevant to overall narrative quality, a story composite score was created summing across six coding categories. A univariate analysis of variance (ANOVA) was conducted for this composite score. The story composite score for the ASD group was significantly lower than that for the TD group, $F(1, 28) = 9.25$, $p = 0.005$, $\eta_p^2 = 0.25$; HFA $M(SD) = 10.9(3.5)$, TD $M(SD) = 14.4(2.8)$.

Story Completeness. Story completeness was quantified by the number of essential plot events a participant mentioned in his/her 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, HFA $M(SD) = 4.4 (1.5)$, TD $M(SD) = 5.2(0.8)$; $F(1, 28) = 3.54$, $p = 0.07$, $\eta_p^2 = 0.11$.

Story Grammar. Story grammar was quantified by the ratio of Initiating Events, Attempts, and Direct Consequences to the total number of T-units. The mean (SD) story grammar ratio in the HFA group was 0.60 (0.23); in the TD group, it was 0.62 (0.20). These scores did not differ between groups, $F(1, 28) = 0.07$, $p = 0.80$, $\eta^2_p = 0.002$.

A second index of overall story goodness was calculated by plotting story completeness against story organization, as shown in Figure 1. In previous research, story goodness is indexed by relatively high scores for both measures (completeness and story grammar), as is the case for results in the second quadrant (upper right quadrant).

Figure 1. Study 1 "Story Goodness" graphs.
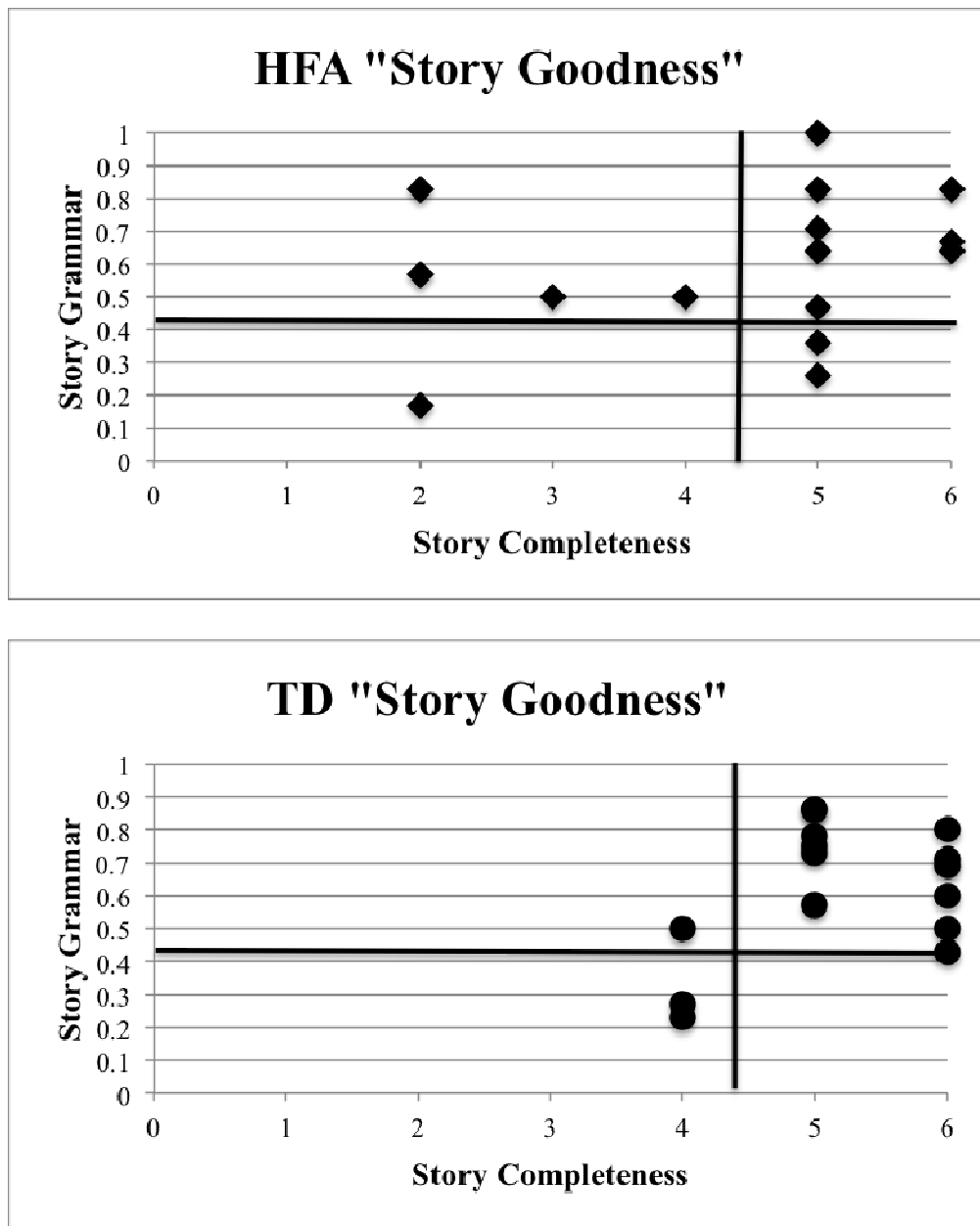
"Story Goodness" was plotted as a function of story grammar and story completeness. Quadrants were defined by cutoff points at 1 standard deviation below the TD mean for both the story grammar and story completeness measures. Quadrant 2 (upper right quadrant) is best representative of a "good" story, as it is high on both story grammar and story completeness. The narrative performance of the HFA group is clearly much more varied than the TD group, especially on the measure of Story Completeness.

Additional Coding Categories. We evaluated Setting, Internal Response, Added/Invented Detail and Total T-units categories separately. These data are also shown in Table 3. While the scores for the HFA group were numerically higher for Added/Invented Detail and numerically lower for all other categories, none of the group differences were significant. Taken together, results indicated that while the HFA group narrations were in general of a lesser quality than those from the TD group, there was no single category of error that drove these group differences; rather, the constellation of a variety of factors seemed to lead to group differences.

Table 4. Study 1 narrative quality ratings.

	HFA M(SD)	TD M(SD)	F	P	η_p^2
Story Goodness	2.46 (0.84)	3.08 (0.88)	3.94	0.06	0.12
Story Cohesiveness	2.90 (0.95)	3.54 (0.71)	4.49	0.05	0.14
Story Accuracy	3.00 (0.85)	3.28 (0.83)	0.84	0.37	0.03
Story Oddness	2.70 (0.85)	2.55 (0.66)	0.30	0.59	0.01

Story Quality Ratings

College undergraduates, naïve to study hypotheses, rated each of 10 narratives to address 4 questions. A repeated measures ANOVA was used to evaluate the story quality ratings, revealing a significant main effect of ratings $F(3, 84) = 3.73$, $p = 0.01$, $\eta_p^2 = 0.12$, a significant main effect of group, $F(1, 28) = 1355.15$, $p < 0.001$, $\eta_p^2 = 0.98$, and no interaction between ratings and group, $F(3, 84) = 1.59$, $p = 0.20$, $\eta_p^2 = 0.05$. Follow-up univariate analyses were conducted, with data shown in Table 4. Of the four rating variables, *Story Cohesiveness* showed a significant effect of group. Group differences in *Story Goodness* ratings reached a trend level

of significance at $p = .06$, with the HFA group receiving lower scores on average than the TD

group in both measures.

Table 5. Study 1 within-group correlational analyses.

	Story Composite	Age	PPVT	Story Goodness	NVIQ
Story Composite	-	**-0.66****	**0.53***	0.4	0.01
Age	-0.13	-	-0.39	-0.16	-0.09
PPVT	-0.31	-0.10	-	**0.74****	0.32
Story Goodness	**0.56***	0.07	-0.22	-	**0.57***
NVIQ	0.36	0.05	0.31	0.38	-

Note: Correlations are presented *above* the diagonal for the HFA group and *below* the diagonal
for the TD group.
**p < 0.01, *p < 0.05, ± p < 0.10

Correlational Analyses

Bivariate correlations were conducted separately for the HFA and TD groups between the

Story Composite scores and the following variables: Age, PPVT standard scores, ratings of *Story*

Goodness, and NVIQ. Results are shown in Table 5. In the HFA group, the *Story Composite*

score was significantly correlated with PPVT score, $r = 0.53$, $p = 0.04$, such that better stories

were told by those with bigger vocabularies. Additionally, in the HFA group, the *Story*

Composite score was significantly correlated with age, $r = -0.66$, $p = 0.01$; however, when FSIQ

is controlled for in this correlation, the results are no longer significant, $r = 0.38$, $p = 0.20$,

indicating that this inverse relationship between age and narrative quality is not meaningful.

Ratings of *Story Goodness* in the HFA group were also significantly correlated with PPVT

scores, $r = 0.74$, $p = 0.002$, which again indicates that raters found that better stories were told

by those with bigger vocabularies. There was also a significant correlation between *Story*

Goodness and NVIQ, $r = 0.57$, $p = 0.03$; however, there was no correlation between *Story*

Composite score and NVIQ in either group (ASD, $r = 0.01$, $p = 0.96$; TD, $r = 0.36$, $p = 0.16$). In

the TD group, the only significant correlation was between *Story Composite* score and ratings of *Story Goodness, r* = 0.56, *p* = 0.03, indicating that both of these metrics measure some of the same aspects of narrative quality in typical developing adolescents. These results suggest that the narratives abilities of the adolescents in the TD group are already well-formed and are no longer correlated with basic language skills, such as vocabulary level. Moreover, these findings indicate that there is "room for improvement" in HFA narrative performance.

Study 1 Discussion

This study evaluated pragmatic language abilities and narrative quality in adolescents with HFA and TD using the "story goodness" coding system, real-world quality ratings and the relationship between these metrics. These measures help to further characterize the pragmatic language abilities observed in individuals with ASD to illuminate their resistance to remediation even in adolescents with HFA who are comparable to their TD peers on other language measures.

A broad measure of story quality (i.e. *Story Composite* score) showed a clear group difference. While the HFA group did not significantly differ from the TD group on the individual components of the *Story Composite* score, the constellation of their narrative abilities as a whole was significantly less than their peers, indicating there is still a deficit in their pragmatic language abilities.

The results of this study indicate that adolescents with HFA tell stories that are likely less complete than their TD peers, consistent with prior research (Landa et al., 1995; Suh et al., 2014). In the current study, the trend towards less complete narratives in the HFA group may be attributed to the working memory demand of the narrative task. Yet, despite the less complete nature of the HFA narratives, there were not group differences in narrative length, suggesting

that the HFA group was not as efficient in their story telling as the TD group. Consistent with

our hypothesis, on the measure of story grammar, the two groups were virtually

indistinguishable. This suggests that the HFA group framed their narratives around prototypical

story events in a manner similar to the TD group, indicating this as a particular area of strength.

The findings suggested that individuals with HFA do not struggle with the same narrative

organization issues during the production of a short narrative from memory, as they do when

producing a longer story while narrating a picture book (Diehl et al., 2006; Suh et al., 2014). This

spared story grammar ability during narrative production may contribute to the impaired story

completeness performance 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 internal response descriptions of thoughts

and feelings that have often been studied in ASD similar to previous findings (Beaumont &

Newcombe, 2006; Capps et al., 2000; Norbury & Bishop, 2009; Tager-Flusberg & Sullivan,

1995). Additionally, while studies of HFA narratives are reported to often 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, as

such a straight-forward plot does not allow for as much creativity, where intrusions of odd or

extraneous details would be more expected.

Correlational analyses revealed an interesting relationship between broad narrative

performance and vocabulary fundamentals in the HFA group. In the HFA group, there was a

significant positive correlation between *Story Composite* scores and PPVT scores, as well as

Story Goodness ratings and PPVT scores, that was not present in the TD group. This suggests

that the TD group has reached a "threshold" of narrative performance that is no longer dependent on vocabulary, which is a more basic language skill; however, the findings in the HFA group suggest that their narrative quality is a function of their vocabulary abilities, indicating that improvement in this domain may lead to improved narrative performance.

Naïve raters were sensitive to differences in narrative quality on broad questions regarding *Story Goodness* and *Story Cohesiveness*. It is particularly interesting that these raters detected group differences just by reading the transcribed narrations, without the influence of interpersonal characteristics such as verbal dysfluency, prosody, or appearance. This suggests that differences in narrative quality are particularly salient in the "real-world", and likely play a role in the day-to-day social interactions of individuals with ASD.

Study Limitations

The current study was limited in several regards. First, the narrative elicitation task involved a greater working memory demand than studies that elicit narrations while participants examine wordless picture books, yet we do not have measures of working memory or executive functioning abilities for this sample. Assessment of executive functioning, including tests of planning, organization, and working memory, would be useful to illuminate how these abilities relate to narrative quality. Furthermore, the inclusion of a control condition where participants are able to examine picture stimuli while narrating a story would allow a more direct comparison of the differing demands of the task with and without the presence of the picture stimuli.

In summary, the findings of the current study are consistent with previous narrative research in ASD in many ways. First, a broad measure of narrative quality, created from summed scores across more specific narrative components, showed that the HFA group 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 their story telling. 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 from naïve raters, which reported lower *Story Goodness* and *Story Cohesiveness* for the HFA group. These findings indicate that it is the combination of several slightly impaired aspects of narrative production that contribute to an overall poorer narrative quality in HFA, rather than specific aspects of narrative performance in isolation.

Study 2

Methods

Participants

Fifteen children and adolescents each with HFA, OO or TD were selected from a larger study of OO such that groups were matched on age and did not differ on IQ. One participant was removed from further analyses due to exceptionally poor experimental task performance, as his Story Completeness and Story Grammar scores were both more than two standard deviations below the mean. Demographic information for the final group of participants is shown in Table 6.

Table 6. Demographic information for participants with HFA, ASD and TD.

	HFA M (SD) Range	OO M (SD) Range	TD M (SD) Range	F	p	Posthoc test
N	14	15	15			
Gender (M:F)	13:1	12:3	14:1			
Age (years)	13.0 (1.6) 10.5-15.7	13.0 (2.15) 9.3-17.1	12.9 (1.5) 9.9 -15.6	0.10	0.99	
WASI						
Nonverbal	106.1 (13.1) 78-120	110.1 (15.2) 87-131	117.8 (12.8) 89-139	2.73	0.08	
Verbal	103.9 (13.9) 81-133	113.0 (18.1) 80-137	111.1 (13.6) 93-136	1.42	0.25	
Full-scale IQ	105.5 (11.0) 80-117	112.8 (15.8) 82-134	115.5 (11.9) 101-142	2.22	0.12	
CELF	102.1 (12.2) 88-124	111.0 (13.1) 79-126	119.6 (8.6) 106-132	9.14	0.001	TD = OO, HFA < TD, p = 0.001, HFA < OO, p = 0.05
PPVT	106.0 (14.2) 86-130	114.5 (14.5) 86-129	122.2 (11.2) 104-143	4.09	0.02	HFA < TD, p = 0.002
ADOS						
Communication	3.3 (1.5) 2-7	0.47 (0.64) 0-2	0.2 (0.6) 0-2	38.93	0.0001	TD = OO, HFA = OO HFA < TD, p < 0.001,
Social	6.8 (2.5) 4-13	0.80 (1.1) 0-3	0.4 (0.8) 0-2	71.05	0.0001	TD = OO, HFA < TD, p < 0.001, HFA < OO, p < 0.001

Participants in the clinical groups were recruited through referrals from private practices and via flyers in the community, autism associations, conference presentations, advertisements in the newspaper and online. Participants in the TD group were recruited through flyers in the community and at the University of Connecticut. For inclusion, all participants were required to have a verbal, nonverbal, and full scale IQ above 77 as measured by the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999).

To be included in the HFA group, participants were required to meet criteria for diagnosis of an ASD on the basis of the ADOS, parent report, and clinical judgment by a clinician with significant ASD experience. To be included in the OO group, participants were required to have a documented diagnosis of ASD by a specialist in the field prior to age 5 and to no longer meet diagnostic criteria for any ASD on the basis of the ADOS and clinical judgment. Additionally, OO participants 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 on the Vineland Adaptive Behavior Scales (Sparrow, Cicchetti & Balla, 2005). Participants were excluded if they exhibited major psychopathology or hearing and vision deficits that would impact their ability to participate in the study. Exclusion criteria also included having a history of seizure disorders, Fragile X disorder, or head trauma with a loss of consciousness. Fein et al. (2013) provides a description of the larger study, including a full description of the recruitment and screening procedures and inclusion criteria.

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 (Lord et al., 2002) was used as the gold standard assessment for verifying and ruling out ASD diagnosis. The SCQ (Rutter et al., 2003) was also used as the parent-report measure of ASD symptomatology, and the PPVT (Dunn & Dunn, 1997) was used as a measure of vocabulary acquisition. Additional measures included:

Wechsler Abbreviated Scales of Intelligence (Wechsler, 1999). The WASI is an abbreviated assessment of cognitive abilities used to obtain a measure of full scale (FSIQ), verbal (VIQ) and nonverbal IQ (NVIQ).

Vineland Adaptive Behavior Scales (VABS; Sparrow et. al, 2005). The VABS is a parent-report measure used to assess adaptive functioning in the domains of Communication, Daily Living Skills, and Socialization.

Clinical Evaluation of Language Fundamentals (CELF, Semel, Wiig, & Secord, 2003). The core subtests of the CELF were administered to evaluate receptive and expressive language functioning of participants. This measure 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 to three sessions.

Participants completed the same experimental task as described in Study 1 with one methodological difference: in Study 1, the picture cards were removed from view before the participants told the story, while in Study 2, the picture cards remained in view during the narration. All narratives were videorecorded for later transcription and analysis. Identical transcription and coding procedures were completed as those described in Study 1.

Reliability. The first author coded all forty-five 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 blind to diagnosis. Interrater reliability was 91% for story completeness and 80% for story grammar in an utterance-by-utterance comparison.

Story Quality Ratings. Forty-seven University of Connecticut undergraduate students rated the quality of the narratives collected from the HFA, OO, and TD participants in return for participation credit in their psychology class. Participants were over 18 years of age and were native speakers of English. Eight participants were excluded due to poor performance on the narrative task.

The participants completed the same rating procedure as described in Study 1; however, participants who evaluated narratives from Study 2 each rated a set of 12 narratives. The order of narratives within these sets was randomized for each participant.

Results

Statistical analyses compared the performance of the HFA, OO, and TD groups on several broad measures of "story goodness" as well as on more specific components of narrative quality. As in Study 1, dependent variables were examined for deviations from the assumptions of normality and sphericity, which were found to be normally distributed. Data were checked for outliers (individuals with scores of more than 2 SD from the group mean). As previously mentioned, one participant was excluded from the HFA group as his narrative performance was more than 2 SD below the mean on both story completeness and story grammar. No other data were excluded from further analyses. Effect sizes were calculated with partial eta squared (η^2_p), which is the proportion of variance attributable to a given effect after partialling out non-error sources of variance (Cohen, 1988).

Multivariate analysis of variance (MANOVA) was again used to evaluate the story coding variables (i.e. *Story Completeness, Story Grammar, Story Composite score, Setting, Internal Response, Added/Invented Detail*, and *Number of T-units*). All four multivariate tests were significant, $F(2, 42) = 795.71$, $p < 0.001$; Pillai's Trace = 0.99, Wilks' Lambda = 0.01, Hotelling's Trace = 159.14. As in Study 1, follow-up univariate analyses of variance (ANOVA) were used to probe for group differences on individual story measures. In Table 7, a summary of coding category data and statistical analyses is provided.

Table 7. Study 2 coding categories as a function of group

|	HFA M(SD)	OO M(SD)	TD M(SD)	F	p	η_p^2	Post-hoc test
Story Completeness	3.9 (1.2)	4.4 (1.7)	4.8 (1.2)	1.75	0.19	0.08	
Story Grammar	0.62 (0.14)	0.60 (0.22)	0.64 (0.13)	0.18	0.84	0.01	
Story Composite Score	10.3 (2.9)	12.4 (4.5)	13.8 (2.7)	3.76	0.03	0.16	HFA < TD, p = 0.002
Coding categories							
Setting	0.43 (0.76)	0.67 (0.62)	1.27 (0.96)	4.36	0.02	0.18	HFA < TD, p = 0.02
Internal Response	0.79 (0.89)	1.00 (0.65)	1.33 (1.04)	1.44	0.25	0.07	
Added/Invented Detail	1.00 (1.30)	1.87 (1.88)	0.87 (1.25)	1.93	0.16	0.09	
Total T-units	8.4 (1.8)	10.9 (4.7)	10.0 (2.3)	2.14	0.13	0.10	

Narrative Analysis

Story Composite Score. A univariate analysis of variance (ANOVA) was conducted for the story composite score. The story composite scores for the HFA, OO, and TD groups differed significantly by group, $F(2, 41) = 3.76$, $p = 0.03$, $\eta^2_p = 0.16$. Post-hoc tests indicated that the HFA/TD group contrast was significant, $F(1, 27) = 11.43$, $p = 0.002$, $\eta^2_p = 0.30$. The HFA/OO contrast missed significance, $F(1, 27) = 2.24$, $p = 0.15$, $\eta^2_p = 0.08$. The OO and TD groups did not differ, $F(1, 28) = 1.07$, $p = 0.31$, $\eta^2_p = 0.04$. Overall story goodness was also examined by plotting story completeness against story organization, as shown in Figure 2.

Story Completeness. The completeness scores, representing the number of key story events expressed in the participant's narrative, were not significantly different between groups, $F(2, 41) = 1.75$, $p = 0.19$ $\eta^2_p = 0.08$.

Story Grammar. The proportion of T-units within story grammar episode structure (i.e. Initiating Events, Attempts, and Direct Consequences) was not significantly different between groups, $F(2, 41) = 0.18$, $p = 0.84$, $\eta^2_p = 0.01$.

Figure 2. Study 2 "Story Goodness" graphs.
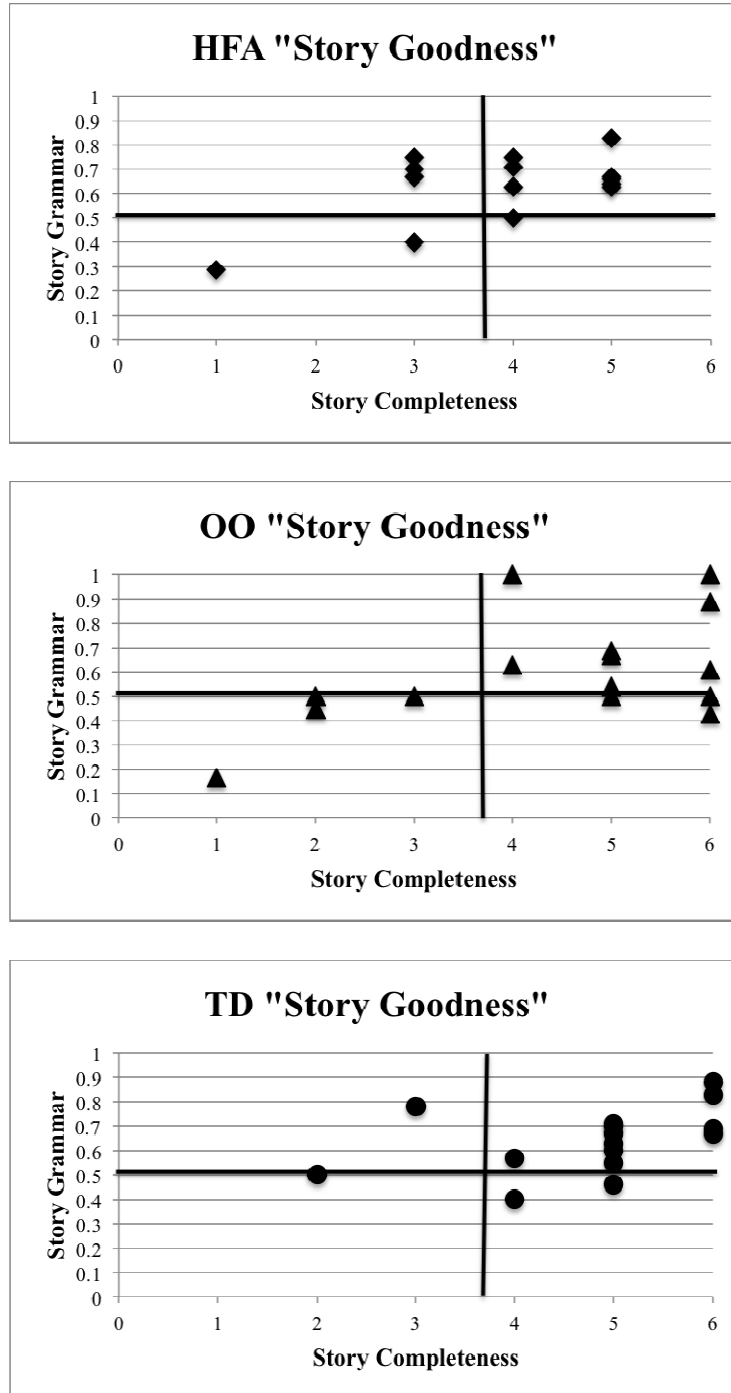
"Story Goodness" was plotted as a function of story grammar and story completeness. Quadrants were defined by cutoff points at 1 standard deviation below the TD mean for both story grammar and story completeness measures. Quadrant 2 (upper right quadrant) is best representative of a "good" story, as it is high on both story grammar and story completeness. The narrative performance of the HFA and OO is more varied than the TD group.

Additional Coding Categories. To follow up on the significant group difference for the *Story Composite* scores, we again evaluated the *Setting* and *Internal Response* categories separately, as well as the categories that were excluded from the *Story Composite* score: *Added/Invented Detail* and *Total T-units*. There was a significant group difference in the *Setting* category, $F(2, 41) = 4.36$, $p = 0.02$, $\eta^2_p = 0.18$. Post-hoc tests indicated that the difference was driven by group differences for HFA and TD groups, $F(1, 27) = 6.75$, $p = 0.02$, $\eta^2_p = 0.20$. No other group comparisons of these coding categories were significant. Results are shown in Table 7.

Table 8. Study 2 narrative ratings as a function of group.

	HFA M(SD)	OO M(SD)	TD M(SD)	F	p	η_p^2	Post-hoc test
Story Goodness	2.15 (0.58)	2.61 (0.89)	3.27 (0.74)	8.1	0.001	0.28	HFA < TD, p < 0.001; OO < TD, p=0.04
Story Cohesiveness	2.71 (0.73)	2.89 (0.89)	3.53 (0.73)	4.47	0.02	0.18	HFA < TD, p = 0.01; OO < TD, p=0.04
Story Accuracy	2.53 (0.68)	2.83 (0.96)	3.47 (0.89)	4.71	0.01	0.19	HFA < TD, p = 0.003
Story Oddness	3.05 (0.55)	2.85 (0.66)	2.49 (0.59)	3.33	0.05	0.14	HFA > TD, p= 0.01

Story Quality Ratings

A repeated measures ANOVA on story quality ratings revealed a significant main effect of group $F(1, 41) = 1860.18$, $p < 0.001$, $\eta^2_p = 0.98$, a main effect of ratings that missed significance, $F(3, 123) = 2.33$, $p = 0.08$, $\eta^2_p = 0.06$, and a significant rating by group interaction, $F(6, 123) = 4.60$, $p < 0.001$, $\eta^2_p = 0.18$. Follow-up univariate analyses were conducted. Results are displayed in Table 8.

Story Goodness ratings for the HFA, OO, and TD groups by participants naïve to study hypotheses showed a significant main effect of group, $F(2, 41) = 8.10$, $p = 0.001$, $\eta^2_p = 0.28$. Post-hoc tests revealed that the TD group had significantly better story goodness ratings than the HFA group, $F(1, 27) = 20.27$, $p < 0.001$, $\eta^2_p = 0.43$, and the OO group, $F(1, 28) = 4.86$, $p = 0.04$, $\eta^2_p = 0.15$. The OO and HFA group ratings did not differ, $F(1, 27) = 2.62$, $p = 0.12$, $\eta^2_p = 0.09$.

Story Cohesiveness ratings showed a significant main effect of group, $F(2, 41) = 4.47$, $p = 0.02$, $\eta^2_p = 0.18$. Post-hoc tests revealed that the TD group ratings were significantly higher than the HFA group, $F(1, 27) = 9.38$, $p = 0.01$ $\eta^2_p = 0.26$, and the OO group, $F(1, 28) = 4.70$, $p = 0.04$, $\eta^2_p = 0.14$. OO and HFA group ratings did not differ, $F(1, 27) = 0.37$, $p = 0.06$, $\eta^2_p = 0.01$.

Story Accuracy ratings showed a significant main effect of group, $F(2, 41) = 4.72$, $p = 0.01$, $\eta^2_p = 0.19$. Post-hoc tests revealed that the narratives from the HFA group were rated as significantly less accurate than those in the TD group, $F(1, 27) = 10.25$, $p = 0.003$, $\eta^2_p = 0.28$. OO group ratings did not differ from those for the HFA group, $F(1, 27) = 0.99$, $p = 0.33$, $\eta^2_p = 0.04$, or TD group, $F(1, 28) = 3.62$, $p = 0.07$, $\eta^2_p = 0.12$.

Story Oddness ratings showed a significant main effect of group, $F(2, 41) = 3.33$, $p = 0.05$, $\eta^2_p = 0.14$. Post-hoc tests revealed that the HFA group ratings were significantly higher than to the TD group, $F(1, 27) = 7.19$, $p = 0.01$, $\eta^2_p = 0.21$; the OO group did not significantly

differ from the HFA group, $F(1, 27) = 0.81$, $p = 0.38$, $\eta^2_p = 0.03$, or TD group, $F(1, 28) = 2.55$, p

$= 0.12$, $\eta^2_p = 0.08$.

Table 9. Within-group correlations for the HFA and OO groups.

	Story Goodness	Age	PPVT	NVIQ
Story Goodness	-	0.37	**0.72****	0.03
Age	0.05	-	0.13	-0.03
PPVT	0.23	0.22	-	-0.21
NVIQ	0.02	0.23	0.51±	-

Note: Correlations are presented *above* the diagonal for the HFA group and *below* the diagonal
for the OO group.
*p < 0.05, ± p < 0.10

Table 10. Within-group correlations for the TD group.

	Story Goodness	Age	PPVT	NVIQ
Story Goodness	-	0.34	0.16	-0.16
Age		-	-0.23	-0.41
PPVT			-	0.33
NVIQ				-

Note: Correlations are presented *above* the diagonal for the TD group.

Correlational Analyses

Bivariate correlations were conducted for the HFA, OO, and TD groups separately

between the *Story Composite* scores and the following variables: Age, PPVT standard scores,

CELF core cores, ratings of *Story Goodness*, and NVIQ. In all three groups, *Story Composite*

scores were not significantly correlated with any variables, except for one notable correlation in

the OO group between the *Story Composite* scores and ratings of *Story Goodness*, $r = 0.62$, $p =$

0.04. To follow-up on these initial finding, the ratings of *Story Goodness* (i.e. the other broad

measure of narrative quality) were compared to the remaining three variables. Results are shown

in Table 9 and Table 10. In the HFA group, ratings of *Story Goodness* was significantly

correlated with PPVT standard scores, $r = 0.72$, $p = 0.004$, but not age, $r = 0.37$, $p = 0.20$, or

NVIQ, $r = 0.03$, $p = 0.92$. In the OO and TD groups, no significant correlations between these variables were observed, with all p's > 0.42 and 0.22, respectively. As in Study 1, these results indicate that the narrative quality of the HFA group may still be dependent on basic language skills as measured by the PPVT.

Study 2 Discussion

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

There were many similarities between Studies 1 and 2 in the comparisons between the HFA and TD groups. As in Study 1, there were no significant group differences between the HFA and TD groups on story grammar, internal response language, added/invented details, and narrative length. Furthermore, the OO group did not differ from either the HFA and TD groups on these measures. Again, this indicates that story structure as measured by story grammar is a particular strength of the clinical groups, similar to findings by Young and colleagues (2005), as other studies found differences in global narrative organization in ASD. The present study also replicated the *Story Composite* score finding seen in Study 1, with the HFA group scoring lower on average on this broad measure of narrative quality than the TD group. Interestingly, the OO group did not significantly differ from either the HFA or TD group.

Contrary to the findings in Study 1, the HFA group did not show a trend towards less complete narratives than the TD group. This is likely due to the decreased working memory demand of the narrative task in the present study, as the picture stimuli were visible during the participants' narrative production.

The HFA, OO, and TD groups performed similarly on most individual measures of narrative quality, with the only one group difference between the HFA and TD groups; overall, the HFA provided significantly less setting information than the TD group. Yet, on the broad measure of narrative quality, the *Story Composite Scores*, the HFA group again received significantly lower scores than the TD group, with the OO group not differing from either group. Similar to Study 1, this finding suggests that is the combination of slightly impaired narrative skills that contributes to diminished narrative quality in HFA, opposed to a single narrative component driving this impairment.

Group differences were observed between the HFA and TD groups across all four ratings domains: *Story Goodness, Story Cohesiveness, Story Accuracy*, and *Story Oddness*. As in Study 1, these findings are particularly robust in that these differences were detectable by naïve raters from reading transcribed narrations, without the influence of other salient factors that characterize the presentation of ASD. Furthermore, the OO and TD groups differed on ratings of *Story Goodness* and *Story Cohesiveness*. This finding supports previous research that suggests that while individuals with OO are virtually indistinguishable from their TD peers in many ways, they may exhibit continued subtle difficulty in pragmatic language (Kelley et al., 2006; Suh et al., 2014). While the story goodness coding system was not able to quantitatively capture group differences between OO and TD, naïve raters detected this difference, implicating the salience of pragmatic language differences in real-world settings. Additionally, graphing "story goodness" as the combination of story grammar and story completeness highlighted the greater variability in narrative performance in the HFA and OO group compared to the TD group. While pragmatic language in HFA should clearly be a focus of intervention, these findings suggest that

adolescents with OO may also benefit from targeted intervention in this particular pragmatic domain.

Limitations

There were several limitations relevant to both Study 1 and Study 2. The narrative samples were relatively brief, as the narrative stimuli were only six cards. These methods differed from many other studies where participants were asked to narrate a longer picture book; therefore, the results of these studies cannot be directly compared to the current studies. However, in many ways, the construction of a short narrative while using working memory to hold and recall events is a more ecologically valid measure of narrative abilities, rather than a page-by-page picture book narration.

<center>**General Discussion**</center>

Overall, the aim of both studies was to use a novel coding system with HFA and OO samples to better characterize their profile of pragmatic language strengths and weaknesses. The results of these studies suggest that working memory abilities may play a role in narrative quality, based on the differences in the narrative completeness observed in Study 1 and Study 2. Across both studies, story grammar or narrative organization abilities were similar in adolescents with HFA, OO, and TD during the production of short narratives. The findings also suggest that a broad measure of narrative quality best captures the deficits in narrative ability observed in HFA, as it is not a single element of narrative production in isolation that drives the differences in narrative quality, but rather the constellation of slight impairments across several components of narrative production. Future research directly comparing the same group of participants on a narrative task with and without the presence of picture stimuli would be help to better tease apart the role of working memory and cognitive load and its impact on narrative quality.

The findings of both studies are beneficial for the assessment and treatment of pragmatic language abilities in higher-functioning individuals with ASD. As many standardized measures do not adequately capture the differences between HFA and TD on this particular aspect of language, narrative elicitation serves as an ecologically valid metric of pragmatic abilities that is salient in the real-world. As suggested by Botting (2002), the use of a parent-report measure of pragmatic language in conjunction with the analysis of an elicited language sample, such as this story goodness analysis, would serve as the best method to characterize pragmatic language abilities. Additionally, as this "story goodness" measure has been previously used in the samples of adults with TBI, future use of this coding procedure across clinical populations and typical development has the potential to inform better treatment strategies across clinical populations and clinical service providers.

In summary, these studies indicate that story grammar is an area of relative strength in HFA, while other aspects of narrative quality, such as completeness, may be a function of working memory demands. Ratings of narrative quality in HFA and TD revealed group differences, indicating that these pragmatic language deficits displayed in narrative production are salient in the real world. Furthermore, consistent with prior research, the detection of differences between the OO and TD groups on ratings of narrative quality suggest that subtle pragmatic deficits may still remain though these adolescents with OO are performing as well as their peers in many other language, social and adaptive domains.

Appendix A. Sample narration.

Story grammar score: 0.88
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)*

Categories:

See Merrit and Liles (1987) for the full coding criteria for the following categories: Initiating Event, Attempt, Direct Consequence, Setting, and Internal Response

Added/Invented Detail – statements that describe story events or excessive details not present in the original story (the picture cards), i.e. additional plans, conflicts, thoughts, wishes, etc.

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. (2000). *Diagnostic and statistical manual of mental disorders IV-TR* (IV-TR ed.). Washington, DC: American Psychiatric Association.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1986). Mechanical, behavioral, 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,* 365-282.

Bishop, D. V. M. (1989). Autism, Asperger's syndrome and semantic-pragmatic disorder: Where are the boundaries*? British Journal of Disorders of Communication, 24*, 107–121.

Bishop, D. (2006). *Children's Communication Checklist—2* U.S. Edition. New York, NY: The Psychological Corporation.

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

Boynton-Hauerwas, L., & 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 and Therapy, 16(*1), 73-86.

Boudreau, D. (2008). Narrative abilities: Advances in research and implications for clinical practice. *Topics in Language Disorders, 28*(2), 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(2), 193–204.

Coelho, C. A. (2002). Story narratives of adults with closed head injury and non-brain-injured

adults: Influence of socioeconomic status, elicitation task, and executive functioning.

Journal of Speech, Language, and Hearing Research, *45*(6), 1232.

Coelho, C. A., Liles, B. Z., & Duffy, R. J. (1995). Impairments of discourse abilities and

executive functions in traumatically brain-injured adults. *Brain Injury*, *9*(5), 471–477.

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

Erlbaum.

Dale, P.S. (1996) Parent report assessment of language and communication. In Cole, K., Dale, P.

and Thal, D., editors, *Assessment of communication and language*, pp. 161–82.

Baltimore, MD: Brookes Publishing.

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*(1), 83–98. doi:10.1007/s10802-005-9003-x

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd ed.). Circle Pines,

MN: American Guidance Services.

Eigsti, I. M., & Bennetto, L. (2009). Grammaticality judgments in autism spectrum disorders:

Deviance or delay. *Journal of Child Language, 19*, 1–23.

Eigsti, I. M., Bennetto, L., & Dadlani, M. (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*(2), 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. *Journal of Child Psychology and Psychiatry 54*(2), 195-205.

Fox, B. J., & Wright, M. (1997). Connecting school and home literacy experiences through cross-age reading. *Reading Teacher*, *50*(5), 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*(3), 267-285.

Houston, G. (1997). The power of story: What I have learned as a writer and a teacher. *Reading Teacher*, *50*(5), 382-395.

Hunt, K. (1965). *Differences in grammatical structures written at three grade levels (NCTE Research Report No. 3)*. Urbana, 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*(6), 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*(2-3), 287-308.

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: Guilford Press.

Landa, R., Martin, M., Minshew, N., & Goldstein, G. (1995). *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. (2011a). Measuring goodness of story narratives. *Journal of Speech, Language, and Hearing Research*, *54*(1), 118.

Lê, K., Coelho, C., Mozeiko, J., Krueger, F., & Grafman, J. (2011b). Measuring goodness of story narratives: Implications for traumatic brain injury. *Aphasiology*, *25*(6-7), 748–760. doi:10.1080/02687038.2010.539696

Liles, B. Z. (1985). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech and Hearing Research, 28*, 123–133.

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*(3), 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, Language and Hearing Research*, *38*(2), 415-425.

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

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2002). *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles: 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*(3), 239–251.

doi:10.1023/A:1024446215446

Loveland, K. A., McEvoy, R. E., & Tunali, B. (1990). Narrative story telling in autism and

Down's syndrome. *British Journal of Developmental Psychology, 8,* 9-23.

MacWhinney, B. (2000). *The CHILDES Project: The database*. Psychology Press.

McDonald, S. (1993). Pragmatic language skills after closed head injury: Ability to meet the

informational needs of the listener. *Brain and Language*, *44*(1), 28–46.

doi:10.1006/brln.1993.1003

Mentis, M., & Prutting, C. A. (1987). Cohesion in the discourse of normal and head-injured

adults. *Journal of Speech, Language and Hearing Research*, *30*, 583-595.

Merritt, D., & Liles, B. (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., Le, 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*(3),

287–313.

Norbury, C. F., Gemmell, T., & Paul, R. (2013). Pragmatics abilities in narrative production: a

cross-disorder comparison. *Journal of Child Language*, 1–26.

doi:10.1017/S030500091300007X

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

Ed.

Roid, G. H. (2003). *Stanford Binet Intelligence Scales, Fifth Edition*.

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

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). San Antonio, TX: Harcourt Assessment, Inc.

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

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*. doi:10.1007/s10803-014-2042-9

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*(03), 241–256. doi:10.1017/S0142716400007281

Thorndyke, P. W., & Yekovitch, F. R. (1980). A critique of schemata as a theory of human story memory. *Poetics, 9*, 23–49.

Tyson, K., Kelley, E., Fein, D., Orinstein, A., Troyb, E., Barton, M., et al. (2013). 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*(3),

648–663. doi:10.1007/s10803-013-1921-9

Troyb, E., Orienstein, A., Tyson, K., Helt, M., Eigsti, I.M., Stevens, M., et al. (2013) Academic

abilities in children and adolescents with a history of autism spectrum disorders who have

achieved optimal outcomes. *Autism.* doi: 10.1177/1362361312473519

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*(3), 204–

212.

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

Psychological Corporation.

Young, E. C., Diehl, J. J., Morris, D., Hyman, S. L., & Bennetto, L. (2005). Pragmatic Language

Disorders in Children with Autism: The use of two formal tests to distinguish affected

children from controls. *Language, Speech, and Hearing Services in Schools*, *36,* 62–72.

.