Narrative Production in Children With Autism Spectrum Disorder and Specific Language Impairment

La production narrative chez des enfants ayant des troubles du spectre de l'autisme ou des troubles spécifiques du langage

KEY WORDS

NARRATIVE PRODUCTION

CROSS-DISORDER COMPARISONS

AUTISM

SPECIFIC LANGUAGE IMPAIRMENT

Paola Colozzo
Heather Morris
Pat Mirenda

Abstract

Purpose: This study examined the narrative productions of a group of verbal children with autism spectrum disorder (ASD) who were compared to two age-matched groups: children with specific language impairment (SLI), and typically developing (TD) peers. The goal was to obtain a profile of narrative abilities of children with ASD along multiple dimensions in order to highlight relevant areas to consider for assessment and intervention.

Method: Three age-matched groups of 6- to 10-year-old children (*N* = 36; *M* age = 102 months) produced two stories from pictures. The analyses considered variables reflecting productivity, content, and form corresponding to vulnerabilities identified in previous studies that have most often considered the narrative abilities of one of these clinical groups, but not both.

Results: The children with ASD were outperformed by their peers with SLI in three areas that ostensibly rely more heavily on perspective-taking abilities: referencing, relevant content, and mental state language. Furthermore, the ASD group produced higher rates of grammatical errors than the SLI group, and there were no differences between these groups for syntactic measures. Not surprisingly, the clinical groups performed below their age-matched TD peers for most measures.

Conclusion: This study adds to the small body of research regarding the narrative abilities of verbal children with ASD. It highlights that, at least for some of these children, structural language should be included as an additional area of focus of assessment and intervention. It supports the use of narrative production tasks as a means of assessing language and communicative abilities across multiple dimensions in order to set intervention goals that aim to improve communicative competence broadly.

Abrégé

Objet : Cette étude a examiné la production narrative chez un groupe d'enfants ayant des troubles du spectre de l'autisme (TSA), s'exprimant à l'oral, comparés à deux groupes d'enfants jumelés selon l'âge : des enfants ayant des troubles spécifiques du langage (TSL), et des pairs ayant un développement typique (DT). L'objectif était d'obtenir un profil des capacités narratives des enfants ayant des TSA selon de multiples dimensions afin de faire ressortir les domaines pertinents à considérer pour l'évaluation et l'intervention.

Méthode : Trois groupes d'enfants de six à dix ans de même âge (*N* = 36; *M* âge = 102 mois) ont produit deux histoires à partir d'images. Les analyses ont considéré des variables qui reflétaient la productivité, le contenu et la forme correspondant aux difficultés identifiées dans des recherches antérieures ayant le plus souvent répertorié les capacités narratives de l'un ou de l'autre de ces groupes cliniques.

Résultats : Les enfants ayant des TSA ont été dépassés dans leur performance par leurs pairs ayant des TSL dans trois domaines qui reposent ostensiblement sur des capacités de prise de perspective : la capacité de faire des références, la pertinence du contenu et le langage exprimant l'état mental. De plus, le groupe TSA a produit un taux plus élevé d'erreurs grammaticales que le groupe TSL et il n'y avait aucune différence entre ces groupes pour les mesures syntaxiques. Tel qu'attendu, les groupes cliniques ont eu une performance inférieure à celle de leurs pairs DT pour la plupart des mesures effectuées.

Conclusion : Cette étude ajoute au petit corpus de recherches concernant les capacités narratives des enfants ayant des TSA, s'exprimant à l'oral. Les résultats mettent l'accent sur le fait que, pour certains de ces enfants, la forme morpho-syntaxique devrait faire partie de l'évaluation et de l'intervention. Ils appuient l'utilisation de tâches de production narrative comme moyen d'évaluer multiples dimensions langagières et communicatives dans le but de fixer des objectifs d'intervention visant à améliorer les compétences en communication.

Paola Colozzo, PhD,
School of Audiology and Speech Sciences, Faculty of Medicine, The University of British Columbia, Vancouver, BC, CANADA

Heather Morris, MSc,
School of Audiology and Speech Sciences, Faculty of Medicine, The University of British Columbia, Vancouver, BC, CANADA

Pat Mirenda, PhD,
Department of Educational and Counselling Psychology, and Special Education, Faculty of Education, The University of British Columbia, Vancouver, BC, CANADA

Narrative production is a demanding task that draws upon linguistic, social, and cognitive abilities (Colozzo, Gillam, Wood, Schnell, & Johnston, 2011; Loveland & Tunali, 1993). The skills required to produce a plot structure and explain the motivations behind actions and events develop well into the school years (Bamberg & Damrad-Frye, 1991; Berman & Slobin, 1994). Furthermore, narrative abilities are a sensitive predictor of later language and literacy outcomes in children with language impairments (Botting, Faragher, Simkin, Knox, & Conti-Ramsden, 2001; Miller et al., 2006; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998).

Narrative production is an assessment context that parallels natural communication events. It can highlight strengths and weaknesses that are not readily observable in standardized testing, and thus provide rich information regarding language in use to guide intervention that aims to improve communicative competence (Botting, 2002; Johnston, 2008). Narratives can be used to obtain language and communicative profiles of children with varied developmental challenges (e.g., Reilly, Losh, Bellugi, & Wulfeck, 2004), although there continues to be limited research regarding the narrative abilities of children and adolescents with developmental disabilities (Finestack, 2012).

Findings from recent research have indicated that a subgroup of verbal children with autism spectrum disorder (ASD) present with language difficulties that overlap with the structural language deficits of children with specific language impairment (SLI) (Kjelgaard & Tager-Flusberg, 2001; Roberts, Rice, & Tager-Flusberg, 2004; Tek, Mesite, Fein, & Naigles, 2014). This has led to considerable debate regarding how such similarities should be interpreted (Ellis Weismer, 2013; Tomblin, 2011). Although "the empirical findings do not offer a clear answer regarding the exact nature of the relationship between ASD and SLI" (Ellis Weismer, 2013, p. 72), cross-disorder comparisons can provide useful information for speech-language pathologists (S-LPs) and other interventionists to support improvements in the language and communication abilities of children with either diagnosis. In particular, such comparisons may point to areas of assessment beyond those that are generally assumed to be vulnerable in one group or the other. Most prior research has considered the narrative abilities of participants with either ASD or SLI but not both groups within a single study. The current study sought to identify profiles of young school-aged children with either diagnosis across multiple dimensions, in comparison to each other as well as to children who are typically developing (TD). The following sections summarize the research that provided a backdrop for the current study.

Narrative Abilities of Children with ASD

Narrative studies that have focused on individuals with ASD are highly heterogeneous with respect to the chronological ages and ability levels of participants, both within and across studies. Participants must, of course, have sufficient spoken expressive language abilities to produce at least a few connected utterances. As a result, most studies have included individuals with ASD who have been described as high-functioning, although this qualifier has been operationalized in different ways, corresponding alternately to overall, nonverbal, or verbal abilities (based on IQ, mental age, or standard scores). For this review, we have dichotomized studies according to whether they included only high-functioning individuals (based on at least one of these criteria) or whether they included less able participants—exclusively or as part of the sample. We have also provided information about key studies in order to describe the samples involved. Whereas studies with high-functioning school-aged children and teenagers with ASD have included age-matched TD peers, those with less able groups have generally compared participants with ASD to language-matched controls with other developmental disabilities or to younger TD participants. Furthermore, because the specifics of the narrative task can have an impact on performance and on the likelihood of finding group differences (e.g., Lai, 2011), the review will focus as much as possible on story generation from picture stimuli.

In terms of productivity and structural language, less able participants with ASD tend to produce stories with fewer words or utterances (or clauses), as well as shorter, syntactically simpler, and less diverse utterances than language-matched controls (Capps, Losh, & Thurber, 2000; Tager-Flusberg, 1995; but see also Tager-Flusberg & Sullivan, 1995). In contrast, results are less consistent when high-functioning participants are compared to age-matched peers. Differences in total words or utterances may (Norbury, Gemmell, & Paul, 2014; Siller, Swanson, Serlin, & Teachworth, 2014) or may not (Norbury & Bishop, 2003; Novogrodsky, 2013; Suh et al., 2014) be present, but results do not seem to pattern according to whether groups with ASD and TD were also matched on a measure of verbal ability. Few studies with high-functioning participants have considered utterance length, but results have been inconsistent even when groups were matched on verbal ability (Norbury et al., 2014; Suh et al., 2014). With respect to syntactic complexity and diversity, most studies have found no differences between high-functioning participants and TD peers with similar language abilities (Diehl, Bennetto, & Young, 2006; Losh & Capps, 2003; Norbury et al., 2014; Novogrodsky, 2013). Norbury and Bishop (2003) did report, however, that participants with

ASD aged 6 to 10 years who had typical nonverbal abilities but depressed scores on expressive and/or receptive language measures, produced fewer complex sentences but more tense marking errors than TD peers.

Two areas have emerged as challenging for participants with ASD across a wide range of age and ability compared to controls: referential cohesion and some aspects of story content. Referential cohesion corresponds to the links between characters, objects, places, or events that appear more than once within a narrative. For instance, the following excerpt illustrates a chain of references to various characters in a story produced by a TD 9-year-old:

> *One day there was a woman and a man.*
>
> *And suddenly a big spaceship came.*
>
> *And these aliens that looked like octopuses on the bottom came.*
>
> *And they had two children and one dog.*
>
> *And then they asked them who they were.*
>
> *They answered that they were aliens from outerspace.*

Compared to language-matched controls, less able participants with ASD produce stories with more imprecise or ambiguous references (e.g., using a pronoun that could refer to more than one character; Loveland, McEvoy, Tunali, & Kelley, 1990). They also do not use linguistic forms in a way that differentiates new versus given information (Tager-Flusberg, 1995), such as introducing a character with a pronoun (e.g., "One morning he woke up") rather than with a full noun phase (e.g., "One morning a boy named Jack woke up"). Moreover, even older high-functioning participants have shown subtle differences in their referential abilities (Norbury et al., 2014; Novogrodsky, 2013; Suh et al., 2014) compared to TD peers, such as higher levels of pronouns where the referent is ambiguous.

Story content and organization has not received as much attention, although a few studies point to this as a likely area of vulnerability. Individuals with ASD across the range of ability include fewer central story components (e.g., problem, actions, and resolution) and/or more repetitive or idiosyncratic material (i.e., information unrelated to the story) in their narratives compared to controls (Diehl et al., 2006; Losh & Capps, 2003; Loveland et al., 1990; Suh et al., 2014; Tager-Flusberg, 1995). Many studies have considered a specific aspect of story content, namely the use of evaluative devices. Evaluations reflect narrators' perspectives regarding "descriptions of mental states or of evaluated outcomes of actions..." (Bamberg & Damrad-Frye, 1991, p. 690). This broad category generally includes mentions of emotional (e.g., *happy, scared*) and cognitive (e.g., *wonder, know*) states (which together comprise frames of mind), character speech, distancing devices that indicate uncertainty (i.e., hedges; e.g., *probably, looks like, kind of*), negative qualifiers, and causal connectors. Group differences with respect to evaluations have not been documented consistently (e.g., Suh et al., 2014; Tager-Flusberg, 1995), possibly because the broad category dilutes differences within and between groups. Various studies do point, however, to a paucity of mental state language (Baron-Cohen, Leslie, & Frith, 1986; Lai, 2011; Pearlman-Avnion & Eviatar, 2002; Tager-Flusberg, 1992).

In summary, unclear referencing, idiosyncratic content, and decreased attention to mental states – all of which rely on theory of mind and perspective-taking abilities (Loveland & Tunali, 1993) – are areas of deficit for many individuals with ASD across a wide range of abilities. Weaknesses in productivity, structural language, and overall story content and organization (i.e., macrostructure) have also emerged, although these aspects have been less studied, and may be more tied to the language and cognitive abilities of participants.

Narrative Abilities of Children with SLI

Many studies have shown that school-aged children with SLI are generally less proficient narrators compared to same-age peers. In addition to length or productivity, most studies have focused either on content or on form, and more recently on how these two aspects may interact.

Form or structural language is an area of particular deficit for many children with SLI (Leonard, 2014). Accordingly, the stories produced by children with SLI tend to be shorter, less grammatically accurate, and comprised of shorter and syntactically less complex utterances than those of TD controls (see Colozzo et al., 2011, for a review). In fact, grammatical accuracy may be an area of particular vulnerability for some children (Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004). Form and content are nonetheless interrelated. Stories with more elaborate content invite more complex syntax, but this can apparently come at a cost, resulting in high levels of grammatical errors (Colozzo et al., 2011).

Regarding references to story characters, children with SLI appear to produce lower levels of appropriate pronominal references or introductions compared to same-age peers (Finestack, Fey, & Catts, 2006; Liles, 1985; Schneider & Hayward, 2010). They may nonetheless demonstrate an ability to adapt to the needs of the listener by providing more precise information in conditions of reduced mutual

knowledge (i.e., depending on whether they had viewed the elicitation material together prior to the telling; Liles, 1985).

Turning to content, the stories told by school-aged children with SLI tend to include fewer main story ideas or story grammar elements (i.e., setting, problem, actions, resolution, internal states) than those produced by TD controls (see Colozzo et al., 2011, for a review). Content may be a relative strength when compared to form for some children, although both dimensions are generally delayed compared to peers (Colozzo et al., 2011; Fey et al., 2004). Some studies point to mental state language as an area of delay in the narratives of children with SLI (Mäkinen, Loukusa, Laukkanen, Leinonen, & Kunnari, 2014; Norbury et al., 2014), although results have been inconsistent (Norbury & Bishop, 2003).

In summary, the accumulated research suggests that children with SLI present weaknesses in multiple dimensions of narrative ability compared to same-age TD peers, with grammatical accuracy and syntactic complexity being particularly vulnerable. Story content, mental states, and referencing may be relative strengths, but differences from controls are nonetheless possible, especially for story macrostructure.

Cross-disorder Comparisons in Narrative Research: ASD and SLI

To our knowledge, only three studies have compared the narrative abilities of children with ASD and SLI. Two studies by Norbury and colleagues have considered the abilities of high-functioning English-speaking children with ASD. In a first study, Norbury and Bishop (2003) analyzed the narrative productions of age-matched groups of children with SLI, ASD, and TD aged 6 to 10 years with typical nonverbal abilities (i.e., scores no more than 1.33 standard deviations below the mean). The two clinical groups did not differ statistically on standardized tests of language ability; in fact, 7 of the 12 children in the ASD group obtained depressed scores on both expressive and receptive language measures. With respect to the narrative task, there were no significant differences between the ASD and SLI groups for productivity, grammatical accuracy, syntactic complexity, referential cohesion, story content, or evaluations. Nonetheless, both clinical groups performed less well than controls on three measures, namely tense marking errors, complex sentences, and ambiguous pronouns.

In a second study, Norbury et al. (2014) included three groups of age-matched children (6 to 15 years) with nonverbal abilities within the normal range (standard scores no more than 1.4 standard deviations below the mean): children with SLI, ASD (with no structural language difficulties; i.e., standardized test scores no more than 1.5 standard deviations below the mean) and TD. Once again, a diversity of measures was considered when analyzing the narratives, but there were few significant group differences between the two clinical samples. The ASD group actually obtained higher scores for internal state language compared to the SLI group. Both clinical groups obtained lower scores than TD peers for mean length of utterance (MLU), although only the SLI group did so for complex syntax. Furthermore, only the ASD group was weaker than controls for story length and referencing. The similarities between the clinical groups and the low productivity and MLU for the group with ASD compared to TD peers are noteworthy given that the participants with ASD scored within the typical range on standardized tests that tapped structural language abilities whereas, by definition, the SLI participants did not. The advantage of the ASD group for mental state language is also perplexing given the profiles of each group. In fact, a study by Ziatas, Durkin, and Pratt (1998) suggests that results could be different for language-matched samples. Eight-year-olds with ASD who completed comprehension and production tasks of belief terms with subtle meaning differences (*know, think,* and *guess*) performed significantly worse than age peers with SLI with similar receptive language (vocabulary and grammar) abilities.

Finally, a study by Manolitsi and Botting (2011) compared the narrative abilities of Greek children with ASD or SLI (aged 4 to 13 years) in a story recall task. The ASD group obtained lower language scores than the SLI group. The groups did not differ in nonverbal abilities, although mean scores were in the low average range (nonverbal IQ, M_{ASD} = 84, range 70-102; M_{SLI} = 87, range 81-92). The ASD group performed less well than the SLI group on two narrative measures – a rating of appropriate content that included both plot structure and character intentions, and a measure of referencing. It is difficult to compare this study with prior research given the coding scheme used. In particular, grammaticality and syntactic complexity were not considered. The findings nonetheless suggest that story macrostructure and referencing may be vulnerable aspects of narrative in individuals with ASD compared to peers with SLI, although the difference in language scores must be kept in mind. Nonetheless, an advantage for children with SLI compared to ASD peers regarding referencing abilities converges with the results of a study by Baltaxe and D'Angiola (1992) that looked at referential cohesion in play/conversational samples produced by groups matched on expressive and receptive language.

In short, studies involving children with either ASD or SLI indicate that referencing, appropriate content, and mental state language may be most vulnerable in participants with ASD, whereas participants with SLI may be mostly challenged by the grammatical and syntactic demands of narratives. Results from the few cross-disorder studies are far from clear-cut, particularly given variability in the characteristics of the clinical samples. Nonetheless, these studies converge in highlighting more similarities than differences. Moreover, the findings of Norbury and colleagues (Norbury & Bishop, 2003; Norbury et al., 2014) point to possible similarities with respect to vulnerabilities in structural language. The current study extends this work by examining the narrative production skills of clinical samples of children with ASD and with SLI aged 6 to 10 years along multiple dimensions in order to provide a broad picture for comparisons.

Method

Participants

Anonymized data from 36 children (three groups of 12 children each with ASD, SLI, and TD) who participated in prior studies were used for the current research. All aspects of the original research projects were reviewed and approved by the Behavioural Research Ethics Board of the University of British Columbia. All participants were recruited from a large geographic area in British Columbia. The sample of children with ASD came from a longitudinal project examining early intervention outcomes of children who fulfilled diagnostic criteria for ASD according to experienced community-based clinicians (see Bopp & Mirenda, 2011, for details regarding identification procedures and characteristics of the larger sample). We included only children who remained in the study after 53 months and were able to complete the three subtests of the Test of Narrative Language (TNL; Gillam & Pearson, 2004) that were administered across the groups (see below), were English monolingual, and had no other reported co-morbidities or diagnoses (e.g., seizure disorder, genetic disorder) other than language delay or impairment. The resulting sample of 12 children with ASD (11 boys, 1 girl) was heterogeneous regarding language ability. In fact, the children generally obtained standard scores on assessments of expressive and receptive language that were bimodally distributed, with five children obtaining scores within the typical range (no more than 1 *SD* below the mean) on four or five of the following measures, whereas another five children scored below the typical range for all five measures: Peabody Picture Vocabulary Test–III (PPVT-III; Dunn & Dunn, 1997); Expressive One-Word Picture Vocabulary Test (EOWPVT; Brownell, 2000); and the Clinical Evaluation of Language Fundamentals–4 (CELF-4; Semel, Wiig, & Secord, 2003) Formulating Sentences, Recalling Sentences, and Concepts and Following Directions subtests. The two other children had heterogeneous profiles, with low scores on two of the five measures. The Vineland Adaptive Behavior Scales (VABS; Sparrow, Balla, & Cicchetti, 1984) were also administered. The range of composite scores corresponding to the mean standard scores from the Communication, Daily living, and Socialization scales was 42 to 93 (*M* = 71), indicating large within-group variability. See Table 1 for details.

Table 1. Scores for Adaptive Behavior and Language for the Group of Children with ASD

Measures	*M*	(*SD*)	Range
VABS mean composite score	71.2	(16.3)	42.3-92.7
CELF, Formulating Sentences	6.8	(4.3)	1-15
CELF, Recalling Sentences	5.5	(3.4)	1-10
CELF, Concepts and Following Directions	5.3	(4.0)	1-12
PPVT-III	84.2	(13.1)	63-111
EOWPVT	92.1	(14.5)	71-116

Note. VABS = Vineland Adaptive Behavior Scales (Sparrow et al., 1984); composite scores correspond to the mean standard scores from the Communication, Daily living, and Socialization scales. CELF = Clinical Evaluation of Language Fundamentals 4 (Semel et al., 2003). PPVT-III = Peabody Picture Vocabulary Test–III (Dunn & Dunn, 1997). EOWPVT = Expressive One-Word Picture Vocabulary Test (Brownell, 2000). CELF-4 subtests, mean standard score = 10, *SD* = 3. All other measures, standardized mean quotient = 100, *SD* = 15.

Each participant with ASD was matched on chronological age with an SLI-TD pair (7 pairs of boys, 5 pairs of girls) who had completed the TNL in the context of another study (Colozzo et al., 2011). S-LPs identified the children with SLI as presenting with persistent oral language difficulties for which they continued to receive intervention. These children were monolingual speakers of English and had no history of intellectual disability, sensory deficits, frank neurological disorder, or any other developmental diagnoses. All children obtained standard scores within or above the normal range (standard scores ≥ 84) on the Test of Nonverbal Intelligence-3 (TONI-3; Brown, Sherbenou, & Johnsen, 1997) and standard scores of 7 or less (i.e., at or below –1 *SD*) on both the Formulated Sentences and Recalling Sentences subtests of the Clinical Evaluation of Language Fundamentals-3 (CELF-3; Semel, Wiig, & Secord, 1995). The children in the control group were native speakers of English with no history of any developmental or academic problems according to teacher and parental reports. All participant triads (*M* age in years;months = 8;6, range 6;7 to 10;2) were matched within 7 months of age or less. Statistical testing using one-way ANOVAs confirmed that the three groups were well-matched on age $F(2, 33) = 0.30$, $p = .97$, and maternal education $F(2, 32) = 1.92$, $p = .16$. See Table 2.

Narrative Tasks

All participants completed three subtests from the TNL, a standardized test used to measure narrative ability. The main data for the current study come from the two narrative production tasks that assess children's ability to generate an original story with visual support: *Late for School* and *Aliens*. The participants first completed a comprehension task, *Shipwreck*, which required them to answer questions after having heard a story corresponding to a five-picture sequence about a girl whose science project is accidently ruined on her way to school. The Shipwreck story provided a model for the subsequent Late for School task, where the child produces a story from five pictures depicting a boy who faces a series of problems that result in his being late for school. For Aliens, the child had to invent a story from a single picture that illustrates two children who witness an alien spaceship landing in a park. Although both production tasks required the narrator to make inferences beyond what was illustrated, the Aliens task was more difficult given that the elicitation picture functioned as a story stem and did not provide an explicit temporal and causal structure. Standard TNL instructions encourage the child to produce a story that is as long and as complete as possible. Additional probes were provided only if the child seemed to lose attention to the task, did not initiate a narrative (e.g., "How does the story start?"), or seemed to end the narrative without signaling that it was complete (e.g., "Is that the end of your story?"). The three tasks were audio-recorded for later scoring according to the TNL guidelines and transcription (see below).

One-way ANOVAs were used to identify any group differences on the three subtests of the TNL. The main effects of group were significant for both production tasks: Late for School, $F(2, 33) = 21.8$, $p < .001$; Aliens, $F(2, 33) = 11.5$, $p < .001$. Based on Games-Howell posthoc tests, both the

Table 2. Demographic Data and Scores on the Subtests of the Test of Narrative Language, by Group

Measures	Groups					
	ASD (*n* = 12)		SLI (*n* = 12)		TD (*n* = 12)	
	M (SD)		*M (SD)*		*M (SD)*	
Age (months)	101.8	(13.4)	101.3	(13.0)	102.7	(12.8)
Maternal education (years)	13.8	(1.3)	13.0	(1.7)	12.8	(1.1)
TNL, Shipwreck (raw; max. 11)	5.8[a]	(4.1)	9.5[b]	(1.3)	10.3[b]	(0.8)
TNL, Late for School (raw; max. 30)	8.0[a]	(3.2)	10.9[a]	(4.0)	18.0[b]	(4.1)
TNL, Aliens (raw; max. 34)	12.4[a]	(6.3)	13.8[a]	(5.4)	22.4[b]	(4.7)

Note. TNL = Test of Narrative Language (Gillam & Pearson, 2004).
For each measure, groups with different superscripts had statistically different mean scores.

ASD and the SLI groups obtained scores significantly below typical peers for Late for School ($ps \leq .001$, $ds \geq 1.75$) and Aliens ($ps \leq .001$, $ds \geq 1.70$), whereas the two clinical groups did not differ (Late for School, $p = .15$, Aliens, $p = .83$). For the comprehension subtest, Shipwreck, the assumption of homogeneity of variance was violated, so we used tests that did not assume equal variances. The Brown-Forsythe test indicated that the main effect of group was significant, $F(2, 14.1) = 10.8$, $p = .001$. Games-Howell posthoc tests indicated that the ASD group obtained significantly lower mean scores than both the TD ($p = .007$, $d = 1.52$) and the SLI ($p = .027$, $d = 1.22$) groups, whereas the SLI and the TD groups did not differ on this measure ($p = .17$). See Table 2.

Transcription of Narrative Texts

We orthographically transcribed the narratives using Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2012) software and conventions. We segmented utterances into communication units (C-units) following the criteria outlined by Loban (1976). C-units consist of main clauses along with any dependent phrase(s) and clause(s). Conjoined clauses containing coordinated conjunctions (*and, or, but*) were split into separate C-units except when the co-referential subject of the second clause was omitted (e.g., "The boy went downstairs and Ø ate breakfast"). We excluded completely unintelligible or abandoned utterances, story closings (e.g. "the end"), and mazes (i.e. false starts, retraces, and within utterance repetitions) from the main story body.

Coding for Structural Language and Story Content

We coded each of the transcripts with regard to clause structure, types(s) of errors, elements of story grammar, and type(s) of mental states.

Clauses. All main clauses and subordinate clauses in the stories were tagged, including nonfinite subordinate clauses (-ing and -ed participles and the base form used as an infinitive; see Huddleston & Pullum, 2005; e.g., "the girl wants[Cl] to go[Cl] to the UFO") and cases of permissible ellipsis of main verbs (e.g., "He said[Cl] 'go [Cl] back to the ship' / And then they did[Cl]" [ellipsis of go]).

Errors. We coded for grammatical, reference, and lexical errors. *Grammatical errors* included (a) omitted words (mostly closed class words such as prepositions, conjunctions, pronouns, auxiliaries, etc.) and omitted bound morphemes (marking plural, tense, person, etc.) in obligatory contexts; (b) incorrect substitutions of closed class words, including prepositions (e.g., "they landed back to Earth"), pronouns (e.g., *him* vs. *he*), or verb forms (e.g., *goed* vs. *went*); (c) unmotivated changes of tense (i.e., moving from past to present, or vice versa; e.g., "the aliens took them away from their parents / so they run all back home); and (d) utterance-level errors, such as omissions of obligatory arguments or incorrect word order.

Reference errors occurred when the identity of a character, object, or location was unrecoverable from the text. Examples included (a) nonspecific character introduction using a definite determiner or a pronoun (e.g., "the boy woked up"; "the boy went to school and told them|his teacher the problem"); (b) use of a pronoun that can refer to more than one character (e.g., "they|children_or_aliens had a dog"); (c) changing of a character's name as the story progresses; and (d) use of demonstratives for places and things (e.g., "they are going there"). *Lexical errors* corresponded to incorrect use of open-class words (e.g., "He drank his breakfast"; "they decide|realize they don't have proof").

Story elements. Following the story grammar system adapted by Merritt and Liles (1987) from the work of Stein and Glenn (1979), a story is potentially comprised of the following elements: (a) *settings*: information about the social and physical context, including story characters, time, and location; (b) *initiating events*: external and internal events that influence and cause a character to respond; (c) *internal responses*: the psychological state that motivates a character to formulate a goal plan; (d) *attempts*: the application of the goal plan actions meant to cause or lead to a resolution; (e) *direct consequences*: the attainment or nonattainment of the character's goal or other changes in the sequence of events caused by a character's actions; and (f) *reactions*: a character's feelings about the attainment or nonattainment of a goal. Working within this framework, we coded each utterance according to the story element(s) it contained. A given utterance could be coded for more than one story element (e.g., an initiating event and an internal response) or for none. We did not consider the overall story organization (i.e., coherence) when judging whether an utterance contributed any story elements. Some utterances were judged to add no additional story elements because they corresponded to extraneous information, contradictory statements, or repetitive content. These *uncoded* utterances (with respect to content) were nonetheless included as part of the story text and to obtain measures of productivity and structural language. The following excerpt from the narrative of a TD 7-year-old presents an example of an uncoded utterance due to repetitive content:

There were these two kids that were around this park.

And they always came here for a few days.

The boy's name was Jack.

And the girl's name was Rachel.

Jack and Rachel both always came to that park [Uncoded: Repetition].

Finally, episode-bridging events that served as both a direct consequence and an initiating event or as both an internal response and a reaction were coded only once, as an initiating event and an internal response, respectively.

Mental States. In the story grammar framework of Stein and Glenn (1979), mental states are coded as internal responses and reactions. We further categorized these story elements into emotional and cognitive states. Emotional states included feelings and conditions (e.g., *scared, worried, tired*). Cognitive states included awareness, intentionality, decisions, desires, and plans (e.g., *want, think, decide, know*).

Measures

We used the SALT program to extract the following data: total numbers of words, C-units, clauses, errors (by type and overall), story elements (by type and overall), and uncoded C-units (i.e., C-units that did not receive any story element codes). Other measures, including mean length of C-unit in words (MLCU-w), clausal density, and error rates were derived from these data (see Results for details).

Reliability

Interrater agreement was based on a randomly selected sample of 25% of participants (i.e., n = 9, 3 per group). A second rater independently scored the three subtests of the TNL. Agreement levels were 98% (range = 91% to 100%) for Shipwreck, 93% (range = 88% to 100%) for Late for School, and 85% (range = 71% to 94%) for Aliens. Given the lower reliability for the Aliens story, a second rater rescored the three TNL subtests for all participants, and a third judge resolved any disagreements.

Based on independent transcription and coding, mean point-by-point interrater reliability was 99% (range = 98% to 100%) for word-level transcription, 97% (range = 94% to 100%) for parsing into C-units, 80% (range = 60% to 100%) for errors, and 94% (range = 85% to 100%) for clause identification. Given the lower reliability for some variables, a second rater verified all transcripts paying particular attention to coding for errors and clauses, and a third judge resolved any disagreements.

Following extensive training and practice, the first and second authors independently coded all narratives for story elements, determined which utterances were left uncoded (i.e., contained no story elements), and came to a final decision. Interrater reliability based on the discrepancies between the two coders (prior to reaching consensus) was 93% (range = 86% to 100%) for all story element types combined, 94% (range = 78% to 100%) for internal responses/reactions specifically, and 93% (range = 80% to 100%) for categorizing uncoded C-units into repetitive or ambiguous utterances.

Results

Mean scores for the TNL Late for School and Aliens tasks suggested that the ASD and the SLI groups did not differ in terms of their narrative production abilities, but that participants in both clinical groups produced stories that were poorer compared to those of the TD peers. These scores, however, reflect various aspects of narrative ability incorporated into the TNL scoring system to produce a composite score for each story. The following analyses based on the story texts considered whether the groups of participants differed in systematic ways on specific measures of storytelling ability.

Analysis Strategy

The general analysis strategy involved a mixed model ANOVA of story (2) by group (3) for each dependent variable. When the main effect of group was significant, we completed posthoc tests comparing each group pair. To control for differences in productivity, we calculated rates or proportions as appropriate. We entered arcsine transformed values into all statistical analyses, but we report marginal means based on untransformed values for ease of interpretation. We report detailed data only for significant effects; in particular, the story-by-group interactions were never significant.[1]

Story Length

All groups produced shorter stories for Late for School compared to Aliens. For both stories, the two clinical groups produced shorter stories than the TD group. Results were similar whether the measure of story length was utterances (C-Units) or words. See Table 3 for details. The mixed model ANOVA for story length in total words resulted in significant main effects of story, $F(1, 33) = 16.5$, $p < .001$, $\eta_p^2 = .33$, and group, $F(2, 33) = 5.21$, $p = .011$, $\eta_p^2 = .24$. Late for School (M = 73.2 words) elicited shorter stories than did Aliens (M = 112.3 words). Posthoc tests indicated that the ASD group (M = 73.0 words, $p = .042$) and the SLI group (M = 76.6 words, $p = .036$) both produced shorter stories

Table 3. Measures of Productivity, Structural Language, and Content, by Group and by Story

Story	Measure	Groups								
		ASD (*n* = 12)			SLI (*n* = 12)			TD (*n* = 12)		
		M (SD)		Range	*M (SD)*		Range	*M (SD)*		Range
Late for School	Total C-units	8.2	(3.1)	4-16	9.8	(3.4)	6-16	13.8	(5.2)	8-26
	Total words	48.8	(20.8)	19-94	63.2	(24.4)	26-99	107.6	(35.8)	64-189
	MLCU in words	5.9	(1.3)	4.8-9.1	6.4	(1.0)	4.3-8.5	8.0	(1.4)	6.0-10.7
	Clauses per C-unit	1.29	(.19)	1.00-1.71	1.40	(0.25)	1.00-2.00	1.72	(0.29)	1.35-2.27
	Errors per word	0.19	(.12)	0.03-0.37	0.08	(0.06)	0.03-0.19	0.03	(0.02)	0.00-0.09
	Story elements	9.4	(3.8)	4-15	15.4	(5.1)	8-22	21.3	(6.5)	13-33
	Proportion mental states	.02	(.04)	.00-.11	.09	(.08)	.00-.27	.10	(.05)	.00-.15
	Proportion uncoded C-units	.21	(.21)	.00-.50	.02	(.06)	.00-.20	.00	(.02)	.00-.06
Aliens	Total C-units	13.5	(10.1)	4-40	13.4	(6.9)	6-29	19.6	(11.8)	7-45
	Total words	97.2	(81.4)	19-307	90.1	(49.1)	40-188	149.8	(84.1)	46-358
	MLCU in words	6.9	(1.5)	3.8-8.8	6.7	(1.2)	4.4-8.5	8.1	(1.8)	5.0-11.5
	Clauses per C-unit	1.18	(0.35)	0.60-1.65	1.38	(0.29)	0.90-2.00	1.62	(0.33)	1.00-2.36
	Errors per word	0.15	(0.09)	0.04-0.30	0.11	(0.08)	0.03-0.30	0.03	(0.03)	0.00-0.09
	Story elements	16.9	(15.6)	2-55	16.3	(10.3)	5-43	25.7	(13.3)	7-59
	Proportion mental states	.12	(.12)	.00-.31	.17	(.11)	.00-.40	.24	(.08)	.14-.42
	Proportion uncoded C-units	.25	(.25)	.00-.67	.15	(.20)	.00-.56	.07	(.09)	.00-.24

than the TD group (*M* = 128.7 words), but the two clinical groups did not differ (*p* = .98).

Structural Language

Utterance length and complexity. We considered MLCU-w and *clausal density*, which corresponds to the mean number of clauses per C-unit, as measures of utterance length and complexity, respectively. The ASD group had slightly lower MLCU-w for Late for School, and fewer clauses per C-unit for both stories compared to the SLI group. Both clinical groups consistently produced stories with shorter utterances and lower clausal density than the TD group. See Table 3.

We completed two mixed model ANOVAs. For MLCU-w, only the main effect of group was significant, $F(2, 33) = 8.10$, $p = .001$, $\eta_p^2 = .33$. Posthoc tests indicated that the ASD group (MLCU = 6.4 words, $p = .010$) and the SLI group (MLCU = 6.5 words, $p = .008$) both produced shorter utterances than the TD group (MLCU = 8.1 words), but the clinical groups did not

differ (p = .95). The results were parallel for clausal density: only the main effect of group was significant, $F(2, 33) = 10.5$, $p < .001$, $\eta_p^2 = .39$. Posthoc tests indicated that the ASD group (M = 1.23 clauses, p = .001) and the SLI group (M = 1.39 clauses, p = .025) had fewer clauses per C-Unit than the TD group (M = 1.67 clauses), but the two clinical groups did not differ (p = .20).

Errors. We obtained length normalized *error rates* by calculating the total number of errors per word for each child. For the combined error types (i.e., grammatical, referential, and lexical), the ASD group had the highest rate, followed by the SLI group and finally the TD group who rarely made errors. This pattern held for both stories. See Table 3.

The mixed model ANOVA for the total error rate indicated only a significant main effect of group: $F(2, 33) = 20.8$, $p < .001$, $\eta_p^2 = .56$. Posthoc tests indicated that both the ASD (M = .17 errors per word; $p < .001$) and the SLI (M = .10 errors per word; $p < .001$) groups had significantly higher error rates than the TD group (M = .03 errors per word). The difference between the two clinical groups did not reach significance despite a strong trend for the ASD group to have a higher error rate than the SLI group, p = .06.

Regarding the specific error types, the patterns for grammatical and reference errors followed that of the combined error rates, and individual level data were consistent with the group differences: grammatical errors per word, M_{ASD} = .12 (SD = .07), M_{SLI} = .07 (.05), M_{TD} = .02 (.01); reference errors per word, M_{ASD} = .04 (.02), M_{SLI} = .02 (.02), M_{TD} = .01 (.01). For grammatical errors, 8 children with ASD obtained rates above the mean for the SLI group. With respect to reference, for all groups the majority of errors were ambiguous references to characters, rather than to objects or places. These errors were largely due to overuse of definite articles or pronouns for new information, and of ambiguous pronouns generally (i.e., more than one possible referent). A majority of the children in the ASD group (n = 7) made reference errors in both stories, but this was the case for only a minority of children in either the SLI (n = 1) or the TD (n = 2) groups, who instead tended to make reference errors only in Aliens (SLI, n = 7; TD, n = 5). Finally, lexical error rates were low for all groups (all Ms ≤ .01 errors per word).

Story Content and Organization

Total story elements. We first considered how many story grammar elements (all types combined) the children included in their stories. All groups produced stories with fewer elements in Late for School than in Aliens, but to varying degrees, with the SLI group showing the smallest difference. The ASD group included fewer story elements than the SLI group only in Late for School, whereas they produced similar numbers of elements for Aliens. Regardless of story, both clinical groups produced fewer elements than the TD group. See Table 3. The mixed model ANOVA for total story elements resulted in significant main effects of story, $F(1, 33) = 5.69$, p = .023, $\eta_p^2 = .15$, and group, $F(2, 33) = 4.71$, p = .016, $\eta_p^2 = .22$. Late for School (M = 15.4) elicited fewer story elements than did Aliens (M = 19.6). Posthoc tests indicated that the ASD group (M = 13.2) produced significantly fewer elements than the TD group (M = 23.5, p = .030). Despite a strong trend for the SLI group (M = 15.8) to have fewer story elements than their TD peers, these groups did not differ significantly on this measure (p = .08), nor did the two clinical groups (p = .71).

Uncoded utterances. We considered the proportions of C-units in each story that did not contribute to the story content (i.e., contained no coded story elements) because they were ambiguous, extraneous, or repetitive. All groups had lower proportions of uncoded utterances for Late for School compared to Aliens, although this difference was least pronounced for the ASD group, for whom one-fifth to one-quarter of utterances on average did not contribute to the story content. For both stories, the ASD group had the highest proportions of uncoded utterances followed by the SLI group and then the TD group. See Table 3.The mixed model ANOVA for the proportion of uncoded C-units resulted in significant main effects of story, $F(1, 33) = 9.00$, p = .005, $\eta_p^2 = .21$, and group, $F(2, 33) = 7.77$, p = .002, $\eta_p^2 = .32$. Late for School (M = .08) had a lower proportion of uncoded utterances than did Aliens (M = .16). Posthoc tests indicated that the ASD group (M = .23) had a significantly higher proportion of uncoded C-units than both the TD group (M = .04, p = .009) and the SLI group (M = .08, p = .045). The SLI and the TD groups did not differ on this measure (p = .63).

We further considered how the uncoded utterances distributed between ambiguous and repetitive utterances. Collapsing across stories and participants, the ASD group produced 88% (n = 46) ambiguous and 12% (n = 6) repetitive utterances; the SLI group, 79% (n = 19) ambiguous and 21% (n = 5) repetitive utterances; the TD group, 38% (n = 8) ambiguous and 62% (n = 18) repetitive utterances.

Distributions of story element categories. We considered how the story elements were distributed by category. Figures 1 and 2 present the mean numbers of story elements produced by each group for each story according to four categories: (a) settings; (b) initiating events/direct consequences; (c) attempts; and (d) internal responses/reactions. In Late for School, all groups produced mostly initiating events/direct consequences, followed by attempts, settings, and lastly by internal responses/

reactions, which were much less frequent than the other three categories, particularly for the ASD group. The distributions were different for Aliens: for all groups, settings stood out as the most frequent category, initiating events/direct consequences and internal responses/reactions were intermediate categories, and attempts were produced the least frequently. Within group variability was high for all element types, but particularly so in Aliens.

Mental states. We calculated the proportion of story elements corresponding to mental states (sum of internal responses and reactions) for each child, by story. All three groups produced a lower proportion of mental states for Late for School than for Aliens. For both stories, the ASD group had the lowest mean proportion of mental states, followed by the SLI and then the TD group. The SLI and TD groups were more similar for Late for School. Individual variability was high, however, particularly for the clinical groups. See Table 3.

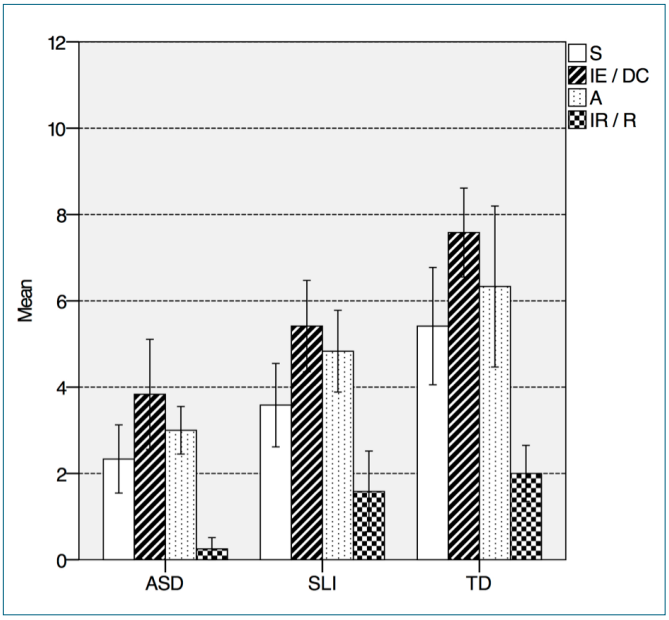

Figure 1. Distribution of Story Elements by Group, Late for School.
S = Settings; IE / DC = Initiating Events and Direct Consequences; A = Attempts; IR / R = Internal Responses and Reactions.
Error bars: 95% CI.

We further considered how the mental state elements distributed between emotional and cognitive states. Collapsing across stories and participants, the ASD group produced 40% (n = 18) emotional and 60% (n = 27) cognitive state elements; the SLI group, 20% (n = 11) emotional and 80% (n = 43) cognitive state elements; the TD group, 25% (n = 24) emotional and 75% (n = 72) cognitive state elements.

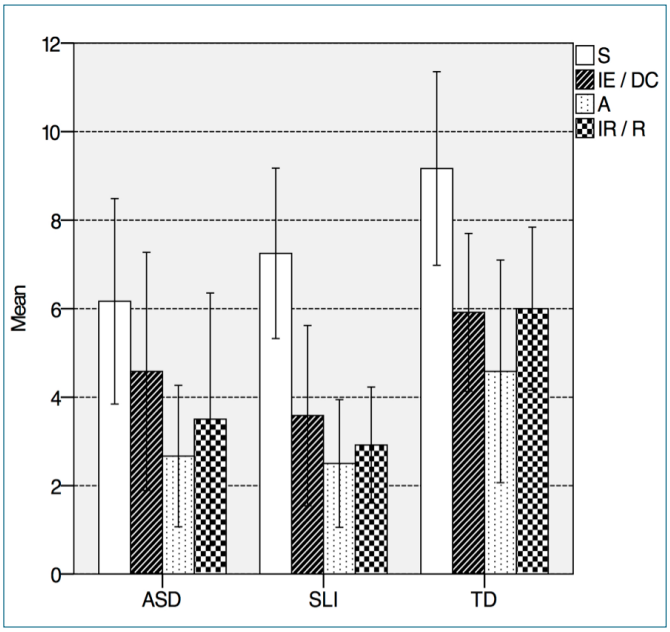

Figure 2. Distribution of Story Elements by Group, Aliens.
S = Settings; IE / DC = Initiating Events and Direct Consequences; A = Attempts; IR / R = Internal Responses and Reactions.
Error bars: 95% CI.

Thus, emotional state elements were less frequent than cognitive state elements for all groups.

Given the high levels of within group variability, we considered individual-level patterns. A minority of children (n = 0 to 3) in each group produced at least one emotional state element per story. The pattern was very different for cognitive state elements, where a minority (n = 2) of children with ASD produced at least one cognitive state element per story, whereas a majority of children (n = 8 or 9) in both the SLI and the TD groups did so. A chi-squared test indicated that there was a significant association between the participant group and whether the children included cognitive elements in their stories, χ^2 (2, N = 36) = 9.59, p = .008, Cramer's V = .52. Fisher's exact tests, two-tailed, confirmed that the two-group comparisons differed significantly between the ASD group and both the TD group (p = .012) and the SLI group (p = .036), to a degree unlikely to be attributable to chance, but that the SLI and TD groups did not differ in this regard (p = 1.00).

Discussion

This study examined the narrative skills of children with ASD and children with SLI who completed two story generation tasks. Narrative production was challenging for both clinical groups who, as expected, obtained lower mean

scores on many dimensions of storytelling ability compared to TD controls. The comparisons of greater interest involve the two clinical groups.

Performance of ASD and SLI Groups Compared to TD Peers

The clinical groups produced, on average, stories with fewer words or utterances, shorter and simpler utterances, and higher levels of errors compared to same-age controls. Thus, both groups consistently performed below TD peers for productivity and structural language. This is clearly in line with prior studies for children with SLI (see Colozzo et al., 2011). It also fits with the limited data from narrative studies involving less able individuals with ASD and language-matched controls (Capps et al., 2000; Tager-Flusberg, 1995), as well as with the results of Norbury and Bishop (2003) regarding tense marking errors and syntactic complexity for a group of children in the same age range as those in the current study who had typical nonverbal abilities but low scores on language measures.

Regarding narrative content measures, only the group with ASD consistently differed from their TD peers. In line with expectations and with previous findings, they produced narratives with fewer central story elements and a higher proportion of utterances that did not contribute to the story (Diehl et al., 2006; Losh & Capps, 2003; Loveland et al., 1990; Suh et al., 2014), as well as a lower proportion of mental states (Baron-Cohen et al., 1986; Lai, 2011; Pearlman-Avnion & Eviatar, 2002; Tager-Flusberg, 1992). The group of children with SLI held an intermediate position for story content, but did not differ significantly from controls, although there was a strong trend for them to produce fewer story elements – as was expected based on prior studies (Colozzo et al., 2011; Merritt & Liles, 1987). That being said, story content may be a relative strength for some children with SLI (Colozzo et al., 2011; Fey et al., 2004). An interesting distinction emerged with respect to utterances that did not contribute to story content. Both clinical groups produced relatively more ambiguous or extraneous utterances, whereas the TD peers produced more utterances with repetitive content. This suggests that TD participants may have been retracing their steps to regain momentum, but were better able to remain within the confines of the developing story.

Only two measures did not differentiate any of the groups. First, only a minority of children in any group mentioned emotional states for both stories. This was unanticipated, given that the visual stimuli represented various character emotions (i.e., frustration, surprise, excitement, fear). It does, however, fit with the scarcity of mental states produced by participants with a range of abilities in past studies (Mäkinen et al., 2014; Siller et al., 2014; Tager-Flusberg & Sullivan, 1995), as well as protracted development of mental state language in the narratives of TD children (Bamberg & Damrad-Frye, 1991). Second, despite differences in productivity, the patterns for the distribution of story element categories were remarkably similar across groups for each story. This suggests that many of the children in the clinical groups were showing an emerging ability to produce stories with an episodic structure. These results are generally in line with those of Merritt and Liles (1987) who found similar hierarchies of story elements for children with SLI and age-matched peers.

Comparisons of the Clinical Groups

Based on scores for the three TNL subtests, the clinical groups were similar in terms of expressive narrative abilities, but the SLI group had an advantage regarding language comprehension, and, in all likelihood, pragmatic abilities. Three of the four areas where the SLI group outperformed the ASD group on the narrative measures presumably rely more heavily on pragmatic skills and perspective-taking abilities: unambiguous reference, relevant content, and mental state language. The other was grammatical accuracy.

Although the few studies that have contrasted participants with ASD and SLI in narrative tasks have not found consistent differences between the clinical groups (Manolitsi & Botting, 2011; Norbury & Bishop, 2003; Norbury et al., 2014), referencing is a well-documented area of weakness in autism research generally (e.g., Loveland et al., 1990; Norbury et al., 2014; Tager-Flusberg, 1995), and two prior studies have found an advantage in this respect for children with SLI compared to peers with ASD (Baltaxe & D'Angiola, 1992; Manolitsi & Botting, 2011). Thus, as anticipated, the ASD group produced more reference errors than did the SLI group; furthermore, only children with ASD tended to make errors in both stories. Appropriate referencing that meets the needs of the listener relies on many skills – including perspective-taking abilities and working memory resources (Arnold, Bennetto, & Diehl, 2009; Whitely & Colozzo, 2013). The children with ASD may have been hampered by their pragmatic deficits relative to participants with SLI, although cognitive differences cannot be ruled out.

Idiosyncratic content was another dimension where the ASD group stood out, as these participants produced more utterances with ambiguous or extraneous information than the SLI group. The high proportions of utterances that did not contribute to the story content (more than one-fifth, on average for each narrative) suggests that the group of children with ASD were more

likely to stray from the stories depicted in the visual stimuli. This result is in accord with prior research that points to this as an area of difficulty for individuals with ASD regardless of ability level (Diehl et al., 2006; Losh & Capps, 2003; Loveland et al., 1990; Suh et al., 2014). The inclusion of more extraneous content could reflect reduced ability to take into account the needs of the listener, limited knowledge of story structure, difficulty remaining on task, limited working memory capacity, or a combination of these factors. Comparatively, the children with SLI produced stories where the intended meaning was more likely to be obscured due to sentence-level organizational difficulties.

There was a strong trend for the ASD group to produce a lower proportion of mental state utterances compared to children with SLI. Furthermore, when we considered individual-level data and honed in on cognitive states, the children with ASD were less likely to produce mentions of cognitive states in each story. Prior research comparing these two clinical groups has not been consistent in this respect (Norbury & Bishop, 2003; Norbury et al., 2014; Ziatas et al., 1998). Nonetheless, results from the current and previous studies (Bamberg & Damrad-Frye, 1991; Norbury & Bishop, 2003; Tager-Flusberg & Sullivan, 1995) converge in suggesting that it may be worthwhile, both in research and clinically, to focus on psychological state terms (or frames of mind) rather than the broad and diverse category of evaluations. Furthermore, within the frames of minds category, it could prove instructive to distinguish between a strict category of cognitive state terms and emotional state terms (Lai, 2011; Mäkinen et al., 2014; Siller et al., 2014). Cognitive states refer to an individual's beliefs, desires, or intentions; thus, they are unobservable, "require an inference of the character's mental state" and "are therefore more indicative of true 'theory of mind' understanding" (Norbury & Bishop, 2003, p. 298). On the other hand, emotional states simply reflect a description of a sensation or an emotion (Astington, 1990). This latter type of frame of mind tends to emerge earlier, and can, to some extent, simply be gleaned from the pictures.

The clinical groups did not differ regarding productivity in words, utterances, or story elements. Regarding story content, it is worth noting that the coding scheme did not take into account the order of story elements, which could have neutralized group differences. Overall coherence and story organization could be the focus of future studies.

The results regarding structural language were arguably the most interesting. The ASD group performed no better than the SLI group on syntactic measures: mean length of utterance and clausal density. Although children with ASD have done less well on syntactic measures relative to language-matched controls in some narrative studies (Capps et al., 2000; Norbury et al., 2014), this has not been a consistent finding (Diehl et al., 2006; Losh & Capps, 2003) – but most of the research has included high-functioning individuals with good expressive language, whereas the current study included a more heterogeneous group of children with ASD with respect to structural language ability.

The ASD group also produced higher rates of grammatical errors than did the SLI group. This was unforeseen, given that grammatical difficulties are a consistent feature of SLI but have not generally been the focus of autism narrative research. This surprising result can nonetheless be explained. Individuals with ASD are a heterogeneous group in terms of structural language abilities, and a sizeable subgroup of children with ASD face similar challenges with structural language as children with SLI (Kjelgaard & Tager-Flusberg, 2001). The fact that at least some of the children with ASD involved in the current study had clinically significant deficits in expressive language and also struggled with morphology and syntax in their narratives is not extraordinary, and is in accord with the findings of Norbury and Bishop (2003) with respect to tense marking difficulties for children with ASD aged 6 to 10 years compared to same-aged TD peers. Still, it was unanticipated that these difficulties would be as pronounced as those exhibited by the group of children with SLI. These results highlight that, at least for some children with ASD, structural language should be included as an additional area of focus of assessment and intervention (Feehan, Francis, Bernhardt, & Colozzo, 2015).

Story Differences

The absence of story by group interactions for measures of productivity, structure, and content reflect that story had a consistent impact across groups. A striking finding was the clear effect of story on the distribution of story elements across categories, and the stability of this effect for all groups. For example, mental states were less frequent in Late for School compared to Aliens. Furthermore, settings was the most frequent element category in the more demanding Aliens story, suggesting that many children were describing the pictures rather than weaving together the elements of a plot structure. Norbury et al. (2014) reported that the groups with autism (with age-expected non-verbal and structural language abilities) and SLI provided more complete information about setting in their stories than did TD controls, with the autism group obtaining the highest ratings on this measure. Such group differences did not emerge here.

The elicitation task can clearly have important effects on variables that tap productivity and narrative content.

Although structural language measures were more stable across stories in the current study, story effects on measures of accuracy and complexity of form are certainly possible. For instance, Colozzo and Whitely (2014) found large story differences in referential cohesion between stories produced by TD 8- to 10-year-olds.

Although there were no significant interactions between group and story for any variable, the ASD and SLI groups performed more similarly for the Aliens story. This story is presumably more demanding given that the elicitation picture does not provide an explicit temporal and causal structure. Moreover, the fantasy theme means that children cannot draw on personal experience for their narrative. Yet, the children with ASD improved in many respects with the second story, whereas those with SLI were more affected by the increased task demands making the two groups appear more similar. In particular, the SLI group's error rate increased, suggesting that these participants were struggling to simultaneously manage content elaboration, syntactic formulation, and grammatical accuracy for the more demanding story (Colozzo et al., 2011).

Clinical Implications

In comparison to their TD peers, the groups with ASD and SLI had similar profiles of narrative abilities along multiple dimensions. Thus, in order to support each child's communicative abilities, clinical assessment should be broad regardless of diagnostic label, and go beyond areas that are generally assumed to be vulnerable for a specific population. Discursive tasks such as narrative production may provide a complementary assessment context to identify areas of difficulty in children with ASD and SLI alike.

Beyond test scores, the analyses of story texts provided much information that would inform intervention. Language sample analysis is a powerful approach to clinical assessment for obtaining descriptive data for goal-setting and measuring progress (Costanza-Smith, 2010; Heilmann, 2010). Clinical tools such as SALT (Systematic Analysis of Language Transcripts; Miller & Iglesias, 2012) provide a means to make this task more efficient. Basic transcription and coding allows clinicians to obtain measures such as total utterances or words, MLU, frequency and types of grammatical errors (i.e., omissions and substitutions). With somewhat more time and effort, it is possible to obtain a measure of clausal density (i.e., subordination index) or to complete the Narrative Scoring Scheme (Heilmann, Miller, Nockerts, & Dunaway, 2010), which is an index that considers story macrostructure, referencing, as well as mental state language. Any of these measures can be compared within a child over time to measure progress —perhaps after a block of therapy or at the end of a school year. Also, under certain conditions, it is possible to compare a child's scores to existing normative databases (see SALT website for details, **www.saltsoftware.com**). Furthermore, additional information can be gleaned by using a combination of elicitation contexts that provide more or less scaffolding, as was the case in the current study.

A story grammar framework such as that provided by Stein and Glenn (1979) can be adapted for clinical use to analyze a child's ability to develop a complete and coherent story and also provide an organizational structure to help reduce processing demands and support comprehension (Johnston, 2008). Clinicians who wish to learn more could consult the examples and detailed coding procedure provided in the appendices by Merritt and Liles (1987) or in the manual that accompanies the Edmonton Narrative Norms Instrument (ENNI; Schneider, Dubé, & Hayward, 2005)—a well-researched narrative assessment tool (normed in Canada), which is freely available at **www.rehabresearch.ualberta.ca/enni**.

Armed with this knowledge, a clinician could easily listen to or read a child's story and identify which types of story elements are present or absent. For example, some children may provide mostly setting information, which could signal that they are in a picture-description rather than a narrative mode; these children would need help learning the key elements of a plot structure. Others may be more advanced, but produce stories that lack psychological causality, and in particular mention of what characters want, think, know, or believe. Helping children to consider and refer to frames of mind—and in particular to unobservable cognitive states—in their personal and fictional narratives could tie in with other goals in the area of social cognition.

Particularly for those children with ASD with more limited linguistic and social-cognitive abilities, these compounding constraints may result in the production of fictional and personal narratives that do not result in successful communicative events given the ambiguity resulting from idiosyncratic content and unclear referencing, combined with simple and ungrammatical form. By helping a child improve in the various areas that underlie narrative ability, the intention is to exert positive change on a child's communicative competence, thereby making successful communicative interactions more likely.

Limitations and Research Implications

This study included a diverse group of children with ASD

that are likely representative of many 6- to 10-years-olds with ASD in S-LP caseloads. There are, nonetheless, undeniable limitations to the current study. The limited sample size calls for caution, and the composition of the samples limits the generalizability to specific groups of children, namely those with cognitive, behavioral, and expressive language profiles that would allow them to complete similar tasks.

The use of archival data also presented some challenges. Although the data succeeded in providing portraits of the participants in each group, we did not have data regarding the nonverbal cognitive abilities of the participants with ASD. Another difference pertained to the administration of the TNL. For the ASD sample, only three subtests of the TNL were completed. This decision was based on concerns regarding the likelihood of participants maintaining attention throughout the protocol. This meant, however, that the children with ASD did not benefit from a model story for the single-picture elicitation task (Aliens); it is thus possible that they could have performed better on measures of story content and, perhaps, productivity. That being said, the data indicated that the scores for the clinical groups on the narrative measures were more similar for the Aliens compared to the Late for School story.

Future studies could replicate and extend the results of this study. Such research with larger samples might consider subgroups within the sample of children with ASD. This could provide relevant clinical information as well as further our understanding regarding the relative constraints of deficits in structural language and social-cognition on discursive abilities that are likely to be associated with both social-communicative and academic success.

Conclusion

This study adds to the limited evidence regarding the narrative abilities of a diverse group of verbal children with ASD aged 6 to 10 years when compared to same-aged peers with typical language and with SLI. From a clinical perspective, it reaffirms that children with SLI and those with ASD would likewise benefit from narrative assessments that consider aspects that have been deemed especially vulnerable in one group or the other, including grammatical accuracy, syntactic complexity, referencing, story grammar, and mental states. In addition to weaknesses in language form that were shared by participants in both clinical groups, the children with ASD were hampered by their more pronounced pragmatic deficits relative to participants with SLI. From a research perspective, this study invites future studies that could tease apart the complex relationships between structural language, social-cognition, and discursive abilities.

References

Arnold, J. E., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition, 110*, 131-146. doi: 10.1016/j.cognition.2008.10.016

Astington, J. W. (1990). Narrative and the child's theory of mind. In B. K. Britton & A. D. Pellegrini (Eds.), *Narrative thought and narrative language* (pp. 151-171). Hillsdale, NJ: Erlbaum.

Baltaxe, C. A. M., & D'Angiola, N. (1992). Cohesion in the discourse interaction of autistic, specifically language-impaired, and normal children. *Journal of Autism and Developmental Disorders, 22*, 1-21. doi: 10.1007/bf01046399

Bamberg, M., & Damrad-Frye, R. (1991). On the ability to provide evaluative comments: Further explorations of children's narrative competencies. *Journal of Child Language, 18*, 689-710. doi: 10.1017/S0305000900011314

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.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.

Bopp, K. D., & Mirenda, P. (2011). Prelinguistic predictors of language development in children with autism spectrum disorders over four-five years. *Journal of Child Language, 38*, 485-503. doi: 10.1017/S0305000910000140

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

Botting, N., Faragher, B., Simkin, Z., Knox, E., & Conti-Ramsden, G. (2001). Predicting pathways of specific language impairment: What differentiates good and poor outcome? *Journal of Child Psychology and Psychiatry, 42*, 1013-1020. doi: 10.1017/S0021963001007843

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Nonverbal Intelligence 3*. Austin, TX: Pro-ed.

Brownell, R. (2000). *Expressive One Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.

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. doi: 10.1023/A:1005126915631

Colozzo, P., Gillam, R. B., Wood, M., Schnell, R. D., & Johnston, J. R. (2011). Content and form in the narratives of children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 54*, 1609-1627. doi: 10.1044/1092-4388(2011/10-0247)

Colozzo, P., & Whitely, C. (2014). Keeping track of characters: Factors affecting referential adequacy in children's narratives. *First Language, 34*, 155-177. doi: 10.1177/0142723714522164

Costanza-Smith, A. (2010). The clinical utility of language samples. *Perspectives on Language Learning and Education*, 17, 9-15. doi: 10.1044/lle17.1.9

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: American Guidance Service.

Ellis Weismer, S. (2013). Developmental language disorders: Challenges and implications of cross-group comparisons. *Folia Phoniatrica et Logopaedica, 65*, 68-77. doi: 10.1159/000353896

Feehan, A., Francis, C., Bernhardt, B. M., & Colozzo, P. (2015). Phonological and morphosyntactic intervention for a twin pair. *Child Language Teaching and Therapy, 31*, 53-69. doi: 10.1177/0265659014536205

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 1301-1318. doi: 10.1044/1092-4388(2004/098)

Field, A. (2009). *Discovering statistics using SPSS (and sex, drugs and rock 'n' roll)* (3rd ed.). Los Angeles, CA: Sage publications.

Finestack, L. H. (2012). Five principles to consider when providing narrative language intervention to children and adolescents with developmental disabilities. *Perspectives on Language Learning and Education, 19*, 147-154. doi: 10.1044/lle19.4.147

Finestack, L. H., Fey, M. E., & Catts, H. W. (2006). Pronominal reference skills of second and fourth grade children with language impairment. *Journal of Communication Disorders, 39*, 232-248. doi: 10.1016/j.jcomdis.2005.12.003

Gillam, R. B., & Pearson, N. A. (2004). *Test of Narrative Language.* Austin, TX: Pro-Ed.

Heilmann, J. J. (2010). Myths and realities of language sample analysis. *Perspectives on Language Learning and Education, 17*, 4-8. doi: 10.1044/lle17.1.4

Heilmann, J. J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*, 154-166. doi: 10.1044/1058-0360(2009/08-0024)

Huddleston, R., & Pullum, G. K. (2005). *A student's introduction to English grammar*. Cambridge, UK: Cambridge University Press.

Johnston, J. R. (2008). Narratives: Twenty-five years later. *Topics in Language Disorders, 28*, 93-98. doi: 10.1097/01.TLD.0000318931.08807.01

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. doi: 10.1080/01690960042000058

Lai, J. Y. (2011). Narrative discourse in school-age children with high-functioning autism (Master's thesis). San Diego State University. Retrieved from http://sdsu-dspace.calstate.edu/handle/10211.10/1281?show=full

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.

Liles, B. Z. (1985). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech and Hearing Research, 28*, 123-133. doi: 10.1044/jshr.2801.123

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.

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. doi: 10.1023/A:3A1024446215446

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. doi: 10.1111/j.2044-835X.1990.tb00818.x

Loveland, K. A., & Tunali, B. (1993). Narrative language in autism and the theory of mind hypothesis: A wider perspective. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 247-266). Oxford, UK: Oxford University Press.

Mäkinen, L., Loukusa, S., Laukkanen, P., Leinonen, E., & Kunnari, S. (2014). Linguistic and pragmatic aspects of narration in Finnish typically developing children and children with specific language impairment. *Clinical Linguistics and Phonetics, 28*, 413-427. doi: 10.3109/02699206.2013.875592

Manolitsi, M., & Botting, N. (2011). Language abilities in children with autism and language impairment: Using narrative as a additional source of clinical information. *Child Language Teaching and Therapy, 27*, 39-55. doi: 10.1177/0265659010369991

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. doi: 10.1044/jshr.3004.539

Miller, J. F., Heilmann, J. J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research and Practice, 21*, 30-43. doi: 10.1111/j.1540-5826.2006.00205.x

Miller, J. F., & Iglesias, A. (2012). Systematic Analysis of Language Transcripts (SALT), Research Version 2012 [Computer Software]. Middleton WI, SALT Software LLC.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders, 38*, 287-313. doi: 0.1080/136820310000108133

Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language, 41*, 485-510. doi: 10.1017/S030500091300007X

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*, 625-632. doi: 10.1007/s10459-010-9222-y

Novogrodsky, R. (2013). Subject pronoun use by children with autism spectrum disorders (ASD). *Clinical Linguistics and Phonetics, 27*, 85-93. doi: 10.3109/02699206.2012.742567

Pearlman-Avnion, S., & Eviatar, Z. (2002). Narrative analysis in developmental social and linguistic pathologies: Dissociation between emotional and informational language use. *Brain and Cognition, 48*, 494-499. doi: 10.1006/brcg.2001.1404

Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language, 88*, 229-247. doi: 10.1016/S0093-934X(03)00101-9

Roberts, J. A., Rice, M. L., & Tager-Flusberg, H. (2004). Tense marking in children with autism. *Applied Psycholinguistics, 25*, 429-448. doi: 10.1017/S0142716404001201

Schneider, P., Dubé, R. V., & Hayward, D. (2005). *The Edmonton Narrative Norms Instrument*. Retrieved from University of Alberta Faculty of Rehabilitation Medicine website http://www.rehabresearch.ualberta.ca/enni

Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language, Speech, and Hearing Services in Schools, 41*, 459-473. doi: 10.1044/0161-1461(2010/09-0040)

Semel, E., Wiig, E. H., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals 3*. San Antonio, TX: The Psychological Corporation.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals 4*. San Antonio, TX: The Psychological Corporation.

Siller, M., Swanson, M. R., Serlin, G., & Teachworth, A. G. (2014). Internal state language in the storybook narratives of children with and without autism spectrum disorder: Investigating relations to theory of mind abilities. *Research in Autism Spectrum Disorders, 8*, 589-596. doi: 10.1016/j.rasd.2014.02.002

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.

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.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41*, 407-418. doi: 10.1044/jslhr.4102.407

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

Tabachnick, B. G., & Fidell, L. S. (2001). *Computer-assisted research design and analysis*. Boston, MA: Allyn and Bacon.

Tager-Flusberg, H. (1992). Autistic children's talk about psychological states: Deficits in the early acquisition of a theory of mind. *Child Development, 63*, 161-172. doi: 10.1111/j.1467-8624.1992.tb03604.x

Tager-Flusberg, H. (1995). 'Once upon a ribbit': Stories narrated by autistic children. *British Journal of Developmental Psychology, 13*, 45-59. doi: 10.1111/j.2044-835X.1995.tb00663.x

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

Tek, S., Mesite, L., Fein, D., & Naigles, L. (2014). Longitudinal analyses of expressive language development reveal two distinct language profiles among young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 44, 75-89. doi: 10.1007/s10803-013-1853-4

Tomblin, B. (2011). Co-morbidity of autism and SLI: Kinds, kin and complexity. *International Journal of Language and Communication Disorders, 46*, 127-137. doi: 10.1111/j.1460-6984.2011.00017.x

Whitely, C., & Colozzo, P. (2013). Who's who? Memory updating and character reference in children's narratives. *Journal of Speech, Language, and Hearing Research, 56*, 1625-1636. doi: 10.1044/1092-4388(2013/12-0176)

Ziatas, K., Durkin, K., & Pratt, C. (1998). Belief term development in children with autism, Asperger syndrome, specific language impairment, and normal development: Links to theory of mind development. *Journal of Child Psychology and Psychiatry, 39*, 755-763.

End Notes

[1] Although ANOVA is robust to violations of normality and with small samples (Norman, 2010), we performed preliminary analyses to test if the assumption of homogeneity of variances was met and to identify outliers. Any data point with a standardized residual value greater than |2.58| was considered an outlier, which corresponds to a 99% confidence interval and is a generally accepted cut-off point for small samples (Tabachnick & Fidell, 2001). Sensitivity analyses were completed with and without outliers, as warranted, and there was no change for significant effects (i.e., they remained significant). To guard against Type 2 error, we used a critical *p*-value of .05 as the criterion for statistical significance throughout the study, but chose the conservative Games-Howell posthoc test, which is recommended in cases when equal variances cannot be assumed (Field, 2009). The assumption of homogeneity of variance was in fact violated twice: for error rate and proportion of uncoded utterances, but only in Late for School.

Acknowledgements

This research was supported in part by funding provided by the British Columbia Ministry of Children and Family Development. We are grateful to the children and families who participated, and thank the many evaluators and research assistants for their work on this project. Portions of this work were presented at the 12th International Congress for the Study of Child Language, held in Montreal, Canada, in July 2011.

Authors' Note

Correspondence concerning this article should be addressed to Paola Colozzo, PhD, School of Audiology and Speech Sciences, The University of British Columbia, 2177 Wesbrook Mall, Vancouver, BC, V6T 1Z3 CANADA Email: paola.colozzo@audiospeech.ubc.ca.