



Published in final edited form as:

*J Speech Lang Hear Res.* 2014 December 1; 57(6): 2221–2233. doi:10.1044/2014\_JSLHR-L-13-0330.

## Use of the ADOS for Assessing Spontaneous Expressive Language in Young Children with ASD: A Comparison of Sampling Contexts

Sara T. Kover, Meghan M. Davidson, Heidi A. Sindberg, and Susan Ellis Weismer

Department of Communication Sciences and Disorders and Waisman Center, University of Wisconsin-Madison

### Abstract

**Purpose**—The current study compared the spontaneous expressive language of children with autism spectrum disorder (ASD) across multiple language sampling contexts: the Autism Diagnostic Observation Schedule (ADOS) and play with an examiner or parent.

**Method**—Participants were children with ASD ( $n = 63$ ; 55 males) with a mean age of 45 months ( $SD = 3.94$ ; Range = 37-53). The number of utterances produced, percent intelligibility, number of different words, mean length of utterance, and the number of requests, comments, and instances of turn-taking were calculated for the ADOS, examiner-child play, and parent-child play. Children were categorized into Tager-Flusberg et al.'s (2009) developmental language phases for each context.

**Results**—Effects of sampling context were identified for all variables examined. The ADOS resulted in fewer utterances and lower structural and pragmatic language performance than examiner- and/or parent-child play. Categorization of children into language phases differed across contexts.

**Conclusions**—Use of the ADOS as a language sampling context may lead to underestimating the abilities of young children with ASD relative to play with an examiner or parent. Researchers and clinicians should be aware of context effects, particularly for assessments designed to observe autism symptoms.

---

Language sampling is a sensitive, ecologically valid, and clinically useful method of assessing expressive language in children with language impairment, including those with autism spectrum disorder (ASD; Costanza-Smith, 2010). Refining best practices for assessment of children with ASD relies upon an understanding of the ways in which the collection and analysis of language samples may impact research- and clinically-driven conclusions about individuals with this neurodevelopmental disorder. The current study examined the potential to utilize a single behavioral measure, the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999), for multiple purposes: both to observe behaviors relevant to ASD—the purpose intended by its developers—and as a context for language sampling to assess spontaneous spoken language.

The ADOS is a semi-structured assessment of social interaction, play, restricted and repetitive behaviors, and communication that is appropriate for individuals with a range of language abilities, from nonverbal to verbally fluent. In particular, we were interested in documenting whether the information provided by a language sample elicited with the ADOS differs from that provided by more traditional language sampling contexts, such as examiner-child or parent-child play-based interactions. Comparisons of expressive language performance among multiple language sampling contexts could serve as a foundation to streamline assessment processes, guide treatment research and practice with respect to the use of language samples as outcome measures, and prove to be informative more generally for the study of behavioral phenotypes associated with neurodevelopmental disorders, such as ASD.

## **Language Samples as Measures of Expressive Language in Young Children: Context and Length**

For young children with typical development and young children with language impairments, characteristics of a language sampling context (e.g., conversation partner, selection of toys) can have significant effects on the language produced during a language sample (Dollaghan, Campbell, & Tomlin, 1990; Hansson, Nettelbladt, & Nilholm, 2000). More structured contexts, such as interview-style conversations, have been shown to elicit a greater number of utterances and utterances with greater mean length of utterance (MLU) than less structured contexts, such as play (Evans & Craig, 1992; Southwood & Russell, 2004). Narrative contexts have been shown to elicit longer and more complex utterances than play or conversations (MacLachlan & Chapman, 1988; Westerveld et al., 2004). Furthermore, the effects of language sampling context might differ across populations of individuals with neurodevelopmental disorders for several aspects of expressive language performance including the amount and complexity of the language produced, highlighting the importance of examining context effects in specific neurodevelopmental disorders (Kover, McDuffie, Abbeduto, & Brown, 2012).

In addition to the context from which a language sample is drawn, some research has indicated that the length of a language sample may impact the estimate of a young child's language ability. For 2- to 3-year old children assessed during unstructured parent-child play, Gavin and Giles (1996) found that test-retest reliability was higher for 20-minute than 12-minute language samples and that reliability increased across measures (e.g., MLU) as the number of utterances in the sample increased. Based on these findings, they recommended a sample comprised of 175 utterances to achieve high reliability, although sufficient reliability was demonstrated with 50 utterance samples. Requiring language samples to contain a particular number of utterances resulted in the loss of participants from analyses, with more than half of participants excluded from analyses of language samples of 175 utterances or more. Emphasizing time elapsed rather than number of utterances produced, Heilmann, Nockerts, and Miller (2010), reported that language samples as short as 3 minutes in length provided consistent estimates of language ability for 2- to 13-year old children during conversation and narration, with no significant differences in performance between 1-min, 3-min, and 7-min language sample cuts. Heilmann et al. did not compare

longer sample lengths. The divergence of findings regarding the extent to which performance varies by length of language sample might be accounted for by specifics of the language sampling context or the developmental level of the children (Heilmann et al., 2010). Nonetheless, it is reasonable to conclude that more utterances generally yield higher reliability, but even relatively short language samples may be informative for understanding spoken language ability.

In summary, research on language sampling methodology suggests that differences in context are highly likely to impact conclusions that are made about young children's spontaneous expressive language ability and that sample length should be considered. In the current study, we examined differences among language sampling contexts (i.e., ADOS, examiner-child play, parent-child play) while holding the length of language samples constant (i.e., comparing 15-minute time segments).

## Language Sampling for Children with ASD

What is known about the language abilities of individuals with ASD is based, in part, on the characterization of spoken language abilities taken from language samples. Research on older children and adolescents with ASD, and primarily those who are high-functioning, has utilized structured language sampling contexts, such as narration (e.g., Hogan-Brown, Losh, Martin, & Mueffelman, 2013; Losh & Capps, 2003; Norbury & Bishop, 2003; Tager-Flusberg & Sullivan, 1995). Of these studies, only a subset has compared expressive language performance across contexts or measures for individuals with ASD. For example, Losh and Capps (2003) found that high-functioning children and young adolescents with ASD demonstrated less complex syntax than typically developing peers during personal narratives, but not during narratives told using a wordless picture book. This study showed that the expressive language of older children with ASD with cognitive abilities in the typical range is susceptible to some effects of language sampling context.

Most research on spontaneous spoken language in children with ASD has been based on unstructured or loosely-structured interactions with a mother, including play (e.g., Hale & Tager-Flusberg, 2005; Swensen, Kelley, Fein, & Naigles, 2007; Tager-Flusberg & Calkins, 1990; Tager-Flusberg et al., 1990). Such research on spontaneous expressive language in children with ASD has revealed significant deficits in domains of structural language ability. Eigsti, Bennetto, and Dadlani (2007), for example, identified syntactic delays in children with ASD relative to developmental level based on language samples drawn from a play session with an examiner. Condouris, Meyer, and Tager-Flusberg (2003) found that language abilities assessed during parent-child play were highly correlated with standardized test performance in children with autism across a large age range (ages 4 – 14 years). Despite the concordance among measures, scores from language samples tended to reveal greater language delay than standardized test scores. Based on these findings, it might be expected that overly-structured assessments might overestimate language skills and that semi-structured or unstructured contexts might result in different representations of language abilities in children with ASD. Overall, little is known about the effects of sampling context on the language of young children with ASD.

## Language Sampling in the Spoken Language Benchmarks Framework

The comparison of language sampling contexts in young children with ASD is motivated not only by the implications of differences among sampling contexts for drawing conclusions about the abilities of children with ASD, but also by the recommendations for intervention research on young children with ASD provided by Tager-Flusberg and her expert colleagues (2009). Tager-Flusberg et al. provided justification for several assessment tools for measuring spoken language in children with ASD and defined developmental language phases with accompanying benchmarks. A key component of their recommendations for assessing the expressive language of young children with ASD was the use of multiple sources of information, including parent report, standardized measures, and natural language samples, so as to obtain valid estimates of language and communication abilities. With a developmental emphasis, language phases were proposed with criteria in four language domains (i.e., phonology, vocabulary, grammar, and pragmatics) for each of three levels of ability: First Words, which aligns with 12 - 18 months in typical development, Word Combinations, which aligns with 18 - 30 months in typical development, and Sentences, which aligns with 30 - 48 months in typical development. Together, this set of recommendations (hereafter, the spoken language benchmarks framework) is significant to the field because, if followed, it would directly impact both the outcome measures selected for research on ASD treatment and the ways in which expressive language would be characterized in studies of young children with ASD, with extensions to the way clinicians might monitor progress. Thus, the spoken language benchmarks framework has the potential to alter the quality and comparability of intervention research for young children with ASD. In the current study, we addressed one aspect of the spoken language benchmarks framework: the choice of language sampling context for assessing spoken language.

Collection of a natural language sample is emphasized by Tager-Flusberg and colleagues (2009) as part of a comprehensive language evaluation. Importantly, a child's performance during a language sample is one of the primary assessments used to determine a child's developmental language phase. To achieve a given developmental phase, a child must meet at least one minimum benchmark criterion within each language domain of that phase. For example, a minimum criterion for the Word Combinations phase in the domain of Grammar is an MLU of 1.8 from a natural language sample. Although benchmark criteria are also provided for parent report and direct assessment, language sampling has several advantages. In particular, language samples are ideal for assessing change in abilities in research because they can be used consistently across levels of ability. In contrast, different standardized tests might be used to assess vocabulary skills at different levels of ability. For instance, one direct assessment suggested within the spoken language benchmarks framework is the Mullen Scales of Early Learning (MSEL; Mullen, 1995) for the Vocabulary criterion for First Words, whereas the Expressive One Word Picture Vocabulary Test-Revised (Gardner, 1990) might be used for Word Combinations. Deciphering the effects of treatment is challenging when outcomes are measured with different assessments over time. In addition, many standardized tests provide only omnibus scores for expressive language (e.g., the MSEL), despite the fact that asynchrony in development across language domains would be expected in children with ASD (Tager-Flusberg et al., 2009). Language samples, on the

other hand, provide specific and separable measures of performance across language domains.

Given the weight that language samples carry, the spoken language benchmarks framework included general recommendations regarding procedures for the collection of a natural language sample. In terms of the length of the sample, the suggestion was to obtain 30 minutes of spontaneous language, perhaps through the aggregation of multiple shorter language samples. For stronger estimates of MLU at more advanced developmental levels, it was advised that, “100 spontaneous (nonimitative/echolalic) child utterances,” be obtained (Tager-Flusberg et al., 2009, p. 650). In terms of the context of the natural language sample, Tager-Flusberg and colleagues indicated that it should be chosen based on the goals of the assessment, including whether the sampling context would target interactions with an experimenter or parent. The authors further suggested that a natural language sample could be drawn from other assessments of communication or autism symptoms, such as the ADOS, the Communication and Symbolic Behavior Scales (CSBS; Wetherby & Prizant, 2002), or the Early Social and Communication Scales (ESCS; Mundy, Hogan, & Doehring, 1996).

These recommendations provide a strong framework for the assessment of children with ASD in treatment settings. Although Tager-Flusberg et al. (2009) noted that research is needed to evaluate “...the relative merits of different types of measures for children with ASD,” (p. 651), no research has compared the advantages and disadvantages of various contexts, including the ADOS, for sampling spoken language in young children with ASD. In the present study, we sought to address this gap in research by comparing the expressive language of a well-characterized sample of preschool children with ASD during the ADOS and two other language sampling contexts that are common in child language research: examiner- and parent-child play.

## The ADOS as a Language Sampling Context

According to the spoken language benchmarks framework, the language sampling contexts that provide frequent motivational opportunities for communication are likely to be best for preschool children at the level of First Words, Word Combinations, or Sentences (Tager-Flusberg et al., 2009). For this reason, the ADOS, examiner-child play, and parent-child play are promising contexts for collection of a language sample that are based on activities used widely in research and intervention settings. Nonetheless, there are several reasons to believe that language samples drawn from some contexts, like the ADOS, might lead to a different profile of strengths and weaknesses than others, such as unstructured play. First, the ADOS is a composite of multiple activities, which vary across modules and across sessions, depending on the order in which they are administered (see Table 1). A module (i.e., a predetermined set of activities) is selected by the examiner based on judgment of the child's expressive language ability (i.e., Module 1 for nonverbal to single words, Module 2 for phrase speech, Module 3 for fluent speech). Although the order of activities is standardized, the examiner may deviate to accommodate the needs of a child. Secondly, the ADOS was designed as an observational assessment of ASD to serve as one source of information in support of a diagnosis. Therefore, it is comprised of social presses and

hierarchies of prompts that elicit communication and opportunities for the observation of behaviors relevant to ASD. This format may lead to a more socially-demanding task than play with either an examiner or a parent, with interactions that vary in terms of structure, language demands, and prompting.

Some previous research has reported on the expressive language of children with neurodevelopmental disorders using the ADOS as a language sampling context; however, the majority of these studies have been based on the same sample of participants with fragile X syndrome (e.g., Barnes et al., 2009; Estigarribia, Martin, & Roberts, 2012; Estigarribia, Roberts, Sideris, & Price, 2011; Price et al., 2008; Roberts et al., 2007). With a focus on children with ASD, one study analyzed expressive morphological and syntactic skills assessed during the ADOS in preschool children with ASD who had IQs greater than 85 (Park, Yelland, Taffe, & Gray, 2012). Park and colleagues concluded that preschoolers with ASD evidenced an uneven profile of morphological and syntactic development relative to developmentally delayed and typically developing comparison groups, with some skills (e.g., use of articles) not differing from comparison groups and other skills (e.g., past tense) impaired. These findings differed from those of Eigsti et al. (2007), who utilized an examiner-child play sample rather than the ADOS as a language sampling context and identified weaknesses in all aspects of morphology and syntax examined. Another study analyzed the language produced by children with ASD with IQs greater than 70 during the ADOS to test an automated error coding system for distinguishing ASD from specific language impairment (Morley, Roark, & van Santen, 2013). Although they were successful in their classifications based on the ADOS, this study included only a single language sampling context. Our goal was to provide the first comparison of the expressive language performance of young children with ASD during the ADOS relative to two other language sampling contexts.

## The Current Study

In consideration of the spoken language benchmarks framework, we assessed expressive language in preschool children with ASD in three contexts: the ADOS, play with an examiner, and play with a parent. For the purpose of systematically comparing contexts to understand the effects of language sampling procedures on the conclusions that are drawn about children's spoken language, we addressed the following research questions: (1) How does the expressive language performance of young children with ASD vary across language sampling contexts (i.e., ADOS vs. examiner-child play vs. parent-child play)?, and (2) Does language sampling context impact classification of young children with ASD into the developmental language phases of the spoken language benchmarks framework? We hypothesized that language sampling contexts would result in different performance for all aspects of expressive language examined: amount produced, phonology, vocabulary, grammar, and pragmatics. We also expected that language sampling contexts would result in different categorizations of children into language phases.

## Method

### Participants

Participants were 63 children (55 males) with ASD with a mean age of 45 months ( $SD = 3.94$ , Range = 37 - 53) recruited as part of a larger longitudinal study (Ellis Weismer et al., 2011; Haebig, McDuffie, & Ellis Weismer, 2013a, 2013b; Ray-Subramanian & Ellis Weismer, 2012; Ray-Subramanian, Huai, & Ellis Weismer, 2011; Venker, Eernisse, Saffran, & Ellis Weismer, 2013) examining language development in toddlers and preschoolers with ASD. Participants in the larger longitudinal study were recruited through local early intervention programs, developmental medical clinics, and posted fliers and magazine and newspaper advertisements in the state of Wisconsin. Appropriate IRB approval and written consent was obtained. Participants with known chromosomal abnormalities, cerebral palsy, frank neurological insults, cleft palate, seizure disorder at the time of recruitment, premature birth, twins, and uncorrected hearing or vision impairment were excluded. All participants were English-only speakers. Participants were seen at up to four visits at one year intervals; data for the current study were drawn from the second visit.

The 63 participants with ASD included in the present analyses were selected from the larger project on the basis of examiner- and parent-child play session transcription. In the larger project, a minimum of 30 child vocalizations was set as a prerequisite to having examiner- or parent-child samples transcribed. Of the 117 participants seen for the second visit, 64 had both examiner- and parent-child play sessions transcribed based on this criterion. Of these 64 participants, 63 participants had a video recording of the ADOS from which a language sample could be drawn. Thus, all ADOS sessions from participants who had both an examiner- and parent-child transcript were transcribed (i.e., the 30-utterance heuristic for transcription of play sessions from the larger project was not applied to the ADOS; however, only one of the 63 participants produced fewer than 30 utterances in the ADOS language sample).

### Measures

**Autism diagnosis**—Children received clinical best estimate diagnoses from trained, expert examiners experienced in child development based on all available information, including the toddler research version of the Autism Diagnostic Inventory-Revised (ADI-R; Le Couteur, Rutter, Lord, & DiLavore, 2006; Rutter, Le Couteur, & Lord, 2003) and the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1999) during their first visit. The ADI-R is a semi-structured parent interview, with questions on social interaction, communication and language, and restricted and repetitive behaviors and stereotyped interests. Participants were re-evaluated at the second visit for autism characteristics using the ADOS to confirm ASD.

The ADOS is a semi-structured assessment that uses standard activities to allow the examiner to observe communication, social, and restricted and repetitive behaviors in individuals with ASD. The ADOS examiner was either research reliable or training to reach reliability, with a reliable examiner present. An ADOS module is selected individually for a child from one of four modules based on the child's expressive language and developmental

level. Participants in the current study received Module 1, Module 2, or Module 3 (Module 4 is appropriate for older adolescents and adults). Each module uses a prescribed series of activities to elicit observable behaviors (Lord et al., 2000). For example, during Free Play, the examiner sets out a variety of toys and objects for the child to play with (e.g., book, doll, balls) and, after observing what the child does, the examiner joins the child and may either continue that activity or initiate others. A list of the activities comprising Modules 1, 2, and 3 is shown in Table 1. Calibrated autism severity scores developed by Gotham, Pickles, and Lord (2009) for the purpose of comparing scores across modules and time were calculated and used to indicate autism severity at the second visit. Calibrated severity scores range from 1 to 10. Scores of 1–3 indicate a non-spectrum classification; scores of 4–5 indicate an autism spectrum classification; and scores of 6–10 indicate an autism classification. Two participants in the current sample received a calibrated severity score of 3 on the ADOS, but met other criteria in order to maintain ASD classification. All other participants received scores in the autism spectrum or autism range. Calibrated severity scores and other participant characteristics are presented in Table 2.

**Norm-referenced measures**—Nonverbal cognitive ability was assessed using the Visual Reception subtest of the Mullen Scales of Early Learning (MSEL; Mullen, 1995). This subtest is intended to assess performance in visual discrimination and visual memory, and is less influenced by motor and verbal abilities. The MSEL is normed from birth to 5; 8 (years; months). The internal consistency ( $r > .53$ ), test-retest reliability ( $r = .85$ ), and concurrent and construct validity are good. Additionally, convergent validity for use of the MSEL with children with ASD was recently established (Bishop, Guthrie, Coffing, & Lord, 2011).

In addition to the language samples described below, the Preschool Language Scale, Fourth Edition (PLS-4; Zimmerman, Steiner, & Pond, 2002) was utilized as a standardized assessment to characterize the language abilities of participants for descriptive purposes. The PLS-4 provides an Auditory Comprehension (AC) standard score ( $M = 100$ ,  $SD = 15$ ) and an Expressive Communication (EC) standard score ( $M = 100$ ,  $SD = 15$ ). The PLS-4 is normed from birth to 6; 11. The internal consistency (AC,  $r > .66$ ; EC,  $r > .73$ ) and test-retest reliability (AC,  $r > .83$ , EC,  $r > .82$ ) are good. Use of the PLS-4 with children with ASD was also recently established (Volden et al., 2011).

**Play-based language samples**—Play-based language samples were 15 minutes in length. During the examiner-child language sample, an examiner—one of several female speech-language pathologists—and the participant played with a Fisher-Price dollhouse. Examiners engaged the child in play and tried to limit the extent to which they asked questions. Examiners also attempted to gloss (i.e., repeat) child utterances. For the parent-child play language sample, a parent (75% mothers) and his or her child played with two sets of toys consisting of Mr. Potato Head and a Fisher-Price farm. Parents were told to play with their child as they usually would at home.

## Procedure

Language samples with an examiner were collected as part of a comprehensive speech-language assessment, including the PLS-4. On a separate day, play with the parent, the



ADOS, and the MSEL were completed. All examiners were female; the examiner who completed the examiner-child play session was not the same as the one who administered the ADOS.

**Transcription**—The language produced by the participant and the examiner or parent was transcribed from digital video using Systematic Analysis of Language Transcripts software (SALT; Miller, Andriacchi, & Nockerts, 2011).

**Selection and transcription of ADOS language samples:** Language samples from the ADOS were selected and transcribed such that they could be directly compared to those elicited during the 15-minute play sessions. For the purposes of the current study, we utilized the length of language sample in time elapsed (i.e., 15 minutes) as the basis for equating the language sampling contexts. Comparing language samples of the same length in time allowed us to evaluate the amount of language produced (i.e., total number of utterances in 15 minutes) in terms of efficiency of elicitation. We selected the first 15 minutes of the ADOS to avoid the possibility that warm-up during the assessment would result in inflating the amount or complexity of language produced during the ADOS relative to the play samples, which were 15 minutes in their entirety. Defining language samples using time elapsed was preferred to using a minimum number of utterances because excluding participants who failed to produce a given number of utterances across all three contexts would limit the generalizability of the findings. Furthermore, the number of utterances produced was a dependent variable of interest. The first 15 minutes of the ADOS was also preferred to selecting a subset of ADOS activities to transcribe for every child (e.g., activities that overlap between Modules 1 and 2 or Modules 2 and 3) because no activities overlap across all three modules. Eliminating participants who completed a module that did not include the selected activities would also limit generalizability. Using time elapsed to define ADOS language samples also has limitations, including the fact that not all transcripts resulted in 50 to 100 utterances, which are desirable lengths for reliability of measures such as MLU (Tager-Flusberg et al., 2009; Gavin & Giles, 1996). For purposes beyond detecting context effects, such as drawing nuanced conclusions about the extent or profile of delay, alternate strategies for selection of a language sample segment would likely be preferred.

Exactly 15 minutes of each examiner- and parent-child play session were transcribed beginning when the examiner or parent initiated the sample until the last utterance at or just before 15 minutes from the beginning of the sample. Using the same start and stop criteria, the first 15 minutes of each ADOS administration was transcribed. If the examiner switched modules during administration, the first 15 minutes of the ultimately scored module was transcribed. As noted above, because the ADOS is a semi-structured assessment, any given portion of the administration (e.g., the first 15 minutes) is likely to contain different activities across participants. Each utterance was assigned a code to signify the activity in which it occurred. The average number of utterances across participants from each activity in the 15 minutes of the ADOS that was transcribed is presented in Table 1.

**General transcription conventions:** Standard SALT procedures for transcription were followed, including the transcription of nonverbal utterances and the coding of overlapping

talk, within and between utterance pauses, mazes, omissions, word errors, and utterance errors. Transcription was further guided by a laboratory manual designed to ensure consistency for ADOS and play sessions. This laboratory manual, for example, enumerated frequently named toys that were to be transcribed as a single word (e.g., JACKINTHEBOX) to avoid artificial inflation of MLU.

Utterances were segmented using phonological units (P-units; Miller et al., 2011). Segmenting utterances based upon P-units documents thought completion based on falling or rising intonation and pauses. In cases in which the child produces conjoined or complex sentences, segmenting is also based on the presence of independent and dependent clauses. Dependent clauses remain conjoined; independent clauses are segmented after use of one conjunction. P-units were selected over communication units (C-units; Loban, 1976), which are independent clauses and any of their modifiers, because P-units may be more sensitive to speaker intentions, even in cases in which grammatical ability may be limited, which was likely to be the case given the young age and limited syntactic complexity of the language of young children with ASD (Miller et al., 2011).

A parent was often present during examiner-child interactions; a parent was present during ADOS administration as necessitated for Modules 1 ( $n = 30$ ) and 2 ( $n = 30$ ). Because of the young ages of the participants and parental preferences, the parent was present in two of three Module 3 administrations, contrary to standardized ADOS administration. Participants' utterances to adults in the room other than the primary conversation partner were not excluded. We also did not exclude imitative utterances or utterances that might have been judged as immediate or delayed echolalia. Although this is contrary to recommendations by Tager-Flusberg et al. (2009), our goal was to compare spoken language produced across contexts without systematically excluding aspects of child language from analyses.

**Agreement:** Transcription agreement was calculated to take into account additions, deletions, or changes of morphemes and additions or deletions of utterances. Transcription agreement was completed for 10% of transcripts randomly selected from each context from the larger project. Project-wide play sample agreement was 93% for morphemes and 96% for segmentation agreement. By context for the visit of interest, transcription agreement was 95% for examiner-child play, 96% for parent-child play, and 90% for the ADOS.

**Variables of interest:** Variables of interest were generated using SALT Research Version 2008 software (Miller & Iglesias, 2008). We assessed the amount of language produced with the total number of utterances, which included those that were nonverbal or otherwise incomplete or unintelligible. The other primary dependent variables were defined within the broad dimensions of the spoken language benchmarks framework: phonology (percent intelligible utterances—or number of consonants produced, for participants who were not at least 50% intelligible, described further below), vocabulary (total number of different words; NDW), grammar (mean length of utterance in morphemes; MLU), and pragmatics (communicative functions, described below). All variables were based on all utterances produced, with the exception of MLU, which was based on the analysis set. The analysis set

was defined as complete and intelligible verbal utterances, thereby excluding nonverbal turns, and abandoned, interrupted, or unintelligible utterances.

**Communicative function coding**—After transcription, each child utterance was coded for its communicative function for the purpose of comparing pragmatic skills across contexts and assigning participants to pragmatic language phases. The primary functions coded were (1) requests for objects, action, or information, (2) comments (including labeling), and (3) turn-taking (including initiations, responses, maintenance, questions, and reporting, each requiring more than a single morpheme). In addition to requests or comments, single morpheme utterances could also be coded as social routines (initiation or response; e.g., *bye*), prompted labels (e.g., *bunny*), acknowledgement (e.g., *oh*), protest (e.g., *no*), and affirmation (e.g., *yes*). Nonlinguistic and nonverbal utterances were coded as such and did not contribute to assignment of participants to pragmatic language phases directly, although they were included in total utterance counts. Utterances that were unintelligible or otherwise ambiguous were coded to denote an unknown communicative function and were also excluded from pragmatic language ability analyses. Communicative function coding reliability was calculated between independent coders for 10% of transcripts (i.e., 6 transcripts from each context). Cohen's kappa was .77,  $p = .009$ , which is considered substantial (Landis & Koch, 1977). The number of (1) requests, (2) comments, and (3) turn-taking utterances were the dependent variables compared across contexts for communicative function performance.

**Classification of participants into language phases:** Participants were classified into developmental language phases separately within each domain based solely on their language samples. For the purpose of the present study, we chose not to include information from other assessments, contrary to the spoken language benchmarks framework, because (1) more than half of the language phase criteria are based on language samples, (2) language samples are comparable across participants, levels of ability, and language domains, and, (3) our focus was strictly on the impact of language sampling context on the conclusions drawn about child language performance.

Following the spoken language benchmarks framework, each language phase was defined by a minimum criterion for phonology, vocabulary, grammar, and pragmatics, as presented in tabular form by Tager-Flusberg et al. (2009; p. 648-649). First Words required at least 4 consonants (phonology), 5 types and 20 tokens (vocabulary), and comments and one other function (pragmatics). Note that all participants who fell below Word Combinations for phonology produced at least 4 consonants. The grammar domain does not apply to First Words. Word Combinations required 50% intelligibility (phonology), 30 different words (vocabulary), MLU of 1.8 (grammar), and commenting, requesting, and turn-taking (pragmatics). Sentences required 75% intelligibility (phonology), 92 different words in a sample of at least 65 utterances (vocabulary), MLU of 3.0 (grammar), and commenting, requesting, and turn-taking with at least two full turns on the same topic following an adult utterance (pragmatics; Tager-Flusberg et al., 2009, p. 648-649). Very few participants were categorized into the same developmental language phases across domains. In fact, 53 participants show a mixed-phase profile (i.e., differed in the developmental language phase

to which they were assigned across phonology, vocabulary, grammar, and/or pragmatics) based on performance during the ADOS; 59 and 54 participants showed mixed-phase profiles based on the examiner- and parent-child play sessions, respectively.

### Analysis Strategy

We tested the effects of language sampling context for each dependent variable (i.e., total number of utterances, intelligibility, NDW, MLU, requests, comments, and turn-taking) using separate repeated-measures ANOVAs with Greenhouse-Geisser corrections, followed by planned pairwise comparisons. The denominator for Cohen's  $d$  was calculated such that the correlation between paired observations was taken into account (Cohen, 1969). To test differences across contexts in the developmental language phases to which children were assigned, we used McNemar's test of marginal homogeneity. McNemar's test is used for  $2 \times 2$  tables of related samples (i.e., matched pairs) with dichotomous variables. These tests were conducted on pairs of language sampling contexts (i.e., ADOS vs. examiner-child play, ADOS vs. parent-child play, examiner- vs. parent-child play) for three dichotomous variables (i.e., First Words vs. all other language phases, Word Combinations vs. all other language phases, and Sentences vs. all other language phases) within each of the four domains. Given the large number of comparisons, we controlled family-wise error rate separately for each outcome using the sequentially rejective Holm procedure (i.e., comparing  $p$ -values ordered smallest to largest for each domain sequentially to  $.05/9 = .0055$ ,  $.05/8 = .0063$ , etc.).

## Results

### Comparison of Language Performance across Contexts

Differences in language performance across contexts were significant for each of the dependent variables examined (see Table 3). The total number of utterances produced across contexts differed,  $F(1.90, 117.48) = 73.74$ ,  $p < .001$ , partial  $\eta^2 = .54$ , with all pairwise comparisons significant,  $ps < .001$ . The fewest utterances were produced during the ADOS relative to examiner-child play,  $d = 0.65$ , and parent-child play,  $d = 1.46$ . More utterances were produced during parent-child than examiner-child play,  $d = 0.95$ . Intelligibility differed across contexts,  $F(1.75, 108.78) = 24.57$ ,  $p < .001$ , partial  $\eta^2 = .28$ , such that children were less intelligible during the ADOS than both examiner-child play,  $p < .001$ ,  $d = 0.63$ , and parent-child play sessions,  $p < .001$ ,  $d = 0.83$ , which did not differ,  $p = .541$ . The NDW produced differed across contexts,  $F(1.79, 111.12) = 29.80$ ,  $p < .001$ , partial  $\eta^2 = .33$ . Fewer words were produced during the ADOS than examiner-child,  $p < .001$ ,  $d = 0.76$ , and parent-child play sessions,  $p < .001$ ,  $d = 0.97$ , which did not differ,  $p = .757$ . In terms of average utterance length, MLU in morphemes for the analysis set differed across contexts,  $F(1.86, 115.08) = 12.50$ ,  $p < .001$ , partial  $\eta^2 = .17$ . MLU was highest for examiner-child play relative to both the ADOS,  $p < .001$ ,  $d = 0.51$ , and parent-child play,  $p < .001$ ,  $d = 0.52$ . MLU did not differ in the ADOS and parent-child play,  $p = .891$ .

For pragmatic performance, the number of requests differed across contexts,  $F(1.85, 114.74) = 3.68$ ,  $p = .031$ , partial  $\eta^2 = .06$ . More requests were made during parent-child play than during examiner-child play,  $p = .034$ ,  $d = 0.27$ , or during the ADOS,  $p = .008$ ,  $d = 0.34$ . The

number of requests made in examiner-child play and the ADOS did not differ,  $p = .958$ . Commenting also differed across contexts,  $F(2.00, 123.92) = 16.04$ ,  $p < .001$ , partial  $\eta^2 = .21$ , with fewer comments made during the ADOS than examiner-child play,  $p < .001$ ,  $d = 0.63$ , and parent-child play,  $p < .001$ ,  $d = 0.60$ , which did not differ,  $p = .852$ . Finally, the number of turn-taking utterances differed across contexts,  $F(1.89, 117.06) = 3.15$ ,  $p = .049$ , partial  $\eta^2 = .05$ . More turn-taking by the child occurred during parent-child play than the ADOS,  $p = .028$ ,  $d = 0.28$ . The amount of turn-taking during examiner-child play did not differ from the ADOS,  $p = .206$ , or parent-child play,  $p = .169$ .

### Categorization into Language Phases across Contexts

Differences in categorization of participants in language phases across contexts are shown in Table 4. For Phonology, the frequency of participants categorized as First Words did not differ across contexts,  $ps > .125$ . Categorization of participants into Word Combinations also did not differ between the ADOS and parent-child play,  $p = .078$ , or between the ADOS and examiner-child play,  $p = .100$ . Categorization into Word Combinations occurred at comparable rates for examiner-child and parent-child play,  $p > .999$ . Differences in categorization into Sentences for Phonology failed to reach significance between the ADOS and examiner-child play,  $p = .014$ , but did differ between the ADOS and parent-child-play,  $p = .004$ . Examiner- and parent-child play did not differ,  $p > .999$ .

For Vocabulary, the frequency of participants categorized as First Words differed between the ADOS and examiner-child and parent-child play,  $ps < .001$ , which did not differ from each other,  $p > .999$ . More participants were categorized as First Words in the ADOS than either of the play sessions. Categorization of participants into Word Combinations did not differ,  $ps > .850$ . Categorization into Sentences was less frequent in the ADOS than examiner-child,  $p = .004$ , and parent-child play,  $p = .004$ , which did not differ,  $p > .999$ .

For Grammar, categorization into Word Combinations failed to reach significance when comparing examiner-child play to the ADOS,  $p = .041$ . Categorization into Word Combinations also did not differ between parent-child play and the ADOS,  $p = .302$ , or parent-child play and examiner-child play,  $p = .383$ . Categorization into Sentences did not significantly differ across contexts,  $ps > .370$ .

For the Pragmatics domain, categorization into First Words differed across contexts, between the ADOS and parent-child play,  $p = .002$ . Examiner-child play did not differ from the ADOS,  $p = .180$ , or parent-child play,  $p = .210$ . Differences in categorization into Word Combinations were not significant,  $ps > .130$ ; differences in categorization into Sentences were also not significant,  $ps > .285$ .

In summary, children with ASD tended to be categorized into lower developmental language phases for phonology, vocabulary, and pragmatics based on performance during the ADOS. Participants were less likely to be categorized into Sentences for phonology based on ADOS performance than performance during parent-child play. Participants were more likely to be categorized into First Words and less likely to be categorized into Sentences based on vocabulary performance during the ADOS than play with an examiner or parent. Finally,

participants were more likely to be categorized as First Words based on pragmatics performance during the ADOS than based on play with a parent.

## Discussion

The purpose of the current study was to examine the effects of language sampling context in a heterogeneous sample of young children with ASD on the amount of language produced, in terms of total number of utterances, and on language performance, in terms of phonology, vocabulary, grammar, and pragmatics. In comparing language samples collected during the ADOS, examiner-child play, and parent-child play, we also examined the extent to which categorization of children into developmental language phases according to Tager-Flusberg et al.'s (2009) spoken language benchmarks framework differed across language sampling contexts. We found context effects for every aspect of language examined, with important implications for researchers and clinicians.

### Context Effects on the Amount of Language Produced

In considering the amount of language produced during a 15 minute language sample, the ADOS resulted in fewer total utterances than either play session language sample. This suggests that transcribing 15 minutes of an ADOS administration may not yield as much information about a child's language abilities because of the small number of utterances obtained relative to language samples drawn during more traditional play samples with an examiner or parent of equal length in time. This finding is noteworthy because language sample methods are resource-intensive, making it vital to devote transcription time to those most likely to yield a sufficient number of utterances to be considered representative of a child's language abilities.

In the current study, we chose to transcribe the first 15 minutes of the ADOS for comparison with examiner- and parent-child play. In tracking the activities in which utterances were produced across participants, it was apparent that a large proportion of utterances analyzed during the ADOS occurred during Free Play. This was usually the first activity administered, but also one in which at least some open-ended opportunities for communication were provided. Other activities yielded very few utterances, perhaps because of their short duration (e.g., Response to Name) or because of the nature of the activity having a focus that does not encourage the production of spoken language (e.g., Functional and Symbolic Imitation). Even the activities of the ADOS designed in part to elicit a language sample are unlikely to yield enough language to be considered representative of the child's skills. Consider, for example, the very small number of average utterances produced during the Conversation activity, which has the aims of eliciting social use of language and assessing the ability to respond to conversational leads given by the examiner (see Table 1). Description of a Picture and Telling a Story from a Book, also from Modules 2 and 3, likewise failed to yield many utterances. A major limitation of the ADOS as a language sampling context is likely to be that its primary purpose is not language sample elicitation. This fact leads the examiner, in accordance with reliable administration of the semi-structured measure, to provide different and fewer opportunities for spontaneous expressive language than would be desired for a representative language sample.

## Context Effects on Structural and Pragmatic Language Performance

Beyond the amount of language produced, performance based on the ADOS yielded lower intelligibility, fewer different words, lower MLU, and fewer requests, comments, and instances of turn-taking than the play-based language samples. Although lower frequencies of word roots and communicative functions might be expected given the lower number of utterances produced during the ADOS, these findings are nonetheless informative. Language samples taken from the ADOS may provide a relatively small set of utterances to analyze, thereby yielding low estimates of vocabulary and pragmatic repertoires for young children with ASD, given the length of the transcribed session. Differences among contexts were also identified for the two variables based not on frequencies, but on percentage and average: intelligibility and MLU, respectively. Thus, differences in abilities between the ADOS and other sampling contexts can be detected for a range of language domains in young children with ASD.

These differences in performance across language sampling contexts had implications for the ways in which children were categorized into developmental language phases of the spoken language benchmarks framework. Indeed, classification differences were observed for the domains of phonology, vocabulary, and pragmatics. In every case, performance based on the ADOS was more likely to lead to classification of participants into a lower developmental language phase relative to a play-based context. This can be taken to indicate that basing conclusions about the language of young children with ASD on administration of the ADOS may underestimate abilities for some domains in the spoken language benchmarks framework.

Although we have identified differences in performance across language sampling contexts, we can only speculate as to why these differences emerged. For example, in the domain of phonology, it is possible that examiners and parents glossed (i.e., repeated) child utterances in the play-based contexts such that intelligibility estimates were artificially inflated. It is also possible that intelligibility during the ADOS was negatively impacted by sound-producing toys that masked the child's speech or the child's movement around the room during activities such as Bubble Play. Park et al. (2012) reported intelligibility of 77% for children with ASD during the ADOS, which is a percent of intelligible utterances that more closely mirrors the scores obtained during the play sessions in the current study. These authors also noted that the nature of the ADOS, including the tendency for children to move around, could have contributed to diminished intelligibility. In this light, conclusions about a child's phonology using a language sample may provide more conservative estimates of intelligibility. For the children with ASD for which phonological development is an area of concern (Kjelgaard and Tager-Flusberg, 2001; Rapin, et al., 2009), a direct norm-referenced assessment, such as the Goldman-Fristoe Test of Articulation, Second Edition (Goldman and Fristoe, 2000) may provide important complementary information (Tager-Flusberg et al., 2009).

It is not our intention to claim that examiner- or parent-child play sessions should necessarily be the gold-standard language sampling contexts for treatment research on children with ASD, but rather to highlight potential consequences of the context selected. Indeed, collecting a language sample based on examiner- or parent-child play may lengthen

an assessment protocol and variability in parent behavior may make it difficult to compare language performance across children. It is also possible that the ADOS yields lower estimates of language ability at a single assessment, but that this might be acceptable for assessing change in language abilities across time or in response to treatment under certain circumstances. It is the researcher's or clinician's responsibility to make an educated choice in sampling context and to draw appropriate conclusions from it. The present data may aid these endeavors.

### Strengths, Limitations, and Future Directions

The current study reported on the expressive language abilities of a relatively large sample of verbal young children with ASD from the perspective of a potentially influential framework, presented by Tager-Flusberg et al. (2009). In fact, this framework has begun to be adopted by some researchers (e.g., Paul, Campbell, Gilbert, & Tsiouri, 2013). In addition to differences in performance among language sampling contexts in the domains of phonology, vocabulary, grammar, and pragmatics, we also highlighted the large number of participants with ASD (over 80% of the current sample) who showed mixed-phase profiles. Given the difficulties children with ASD experience with social interaction, it is not unexpected that pragmatics, for example, might lag behind other language domains, such as phonology. Even within language domains, such as syntax, young children with ASD may display skills that differ in extent of delay (e.g., use of negation vs. use of verb phrases; Park et al., 2012). However, here, we document that very few children perform within the same developmental language level regardless of the language sampling context chosen. Although this finding is descriptive in nature, it is a point worth emphasizing to researchers and clinicians who design treatment studies or treatment plans, either with the spoken language benchmarks framework in mind or on the basis of other perspectives.

Participants with ASD in the current study were relatively heterogeneous; however, generalizability of the results is somewhat limited by the initial exclusion of participants in the larger longitudinal study who produced fewer than 30 vocal utterances during the examiner- and/or parent-child play sessions. By excluding participants who did not have 30 utterances for either the examiner- or parent-child play sessions, conclusions about differences in performance across language sampling contexts cannot be extended to children with ASD who might be considered minimally verbal. It is also conceivable that findings would have differed if our comparisons were made between language samples of 100 utterances or more for each context, likely limiting analyses to participants with the strongest spoken language abilities. For many young children with ASD, it is possible that transcription of an entire ADOS administration or 30 minutes of play with an examiner or parent might still result in fewer than 100 utterances, while also increasing the time required for transcription.

In post-hoc analyses, we repeated comparisons among language sampling contexts separately for subgroups of participants with fewer than 100 utterances ( $n = 27$ ) and participants with more than 100 utterances ( $n = 36$ ) produced during the 15-minute ADOS sample. Conclusions about the amount of language produced and structural language ability among contexts were the same as reported for the full sample for both subgroups. For



requests and turn-taking, context effects were significant for neither subgroup. The pattern of performance differed between the subgroups for only one pairwise comparison (ADOS vs. examiner-child play for commenting), which was significant for the subgroup with fewer than 100 utterances (as well as the full sample), but not the subgroup with more than 100 utterances. These subgroup analyses provide some evidence that the preponderance of our findings were unlikely driven by language samples of too few utterances to be considered representative; however, it is possible that pragmatic performance is the most vulnerable to differential effects of context across language sample lengths.

Given this, it may be that longer segments of the ADOS are necessary to obtain reliable estimates of language ability when utilized as a language sampling context. Future research might evaluate the utility of the ADOS as a language sampling context using different strategies to partition it. Rather than the first 15 minutes, alternatives might include random selection of a given number of one-minute time-segments (Heilmann et al., 2010) or transcribing the language produced during particular activities (e.g., make-believe play, conversation)—the strategy used by Park et al. (2012). Park et al. analyzed activities that were shared in common between Modules 2 and 3, yielding at least 100 utterances for upwards of 85% of participants, who were somewhat older and likely more linguistically advanced than those in the current sample. Establishing which ADOS activities are most suitable for characterizing expressive language abilities was beyond the scope of the current study, although we have shown descriptively that more utterances are likely to be drawn from some activities than others (e.g., Free Play, Joint-Interactive Play). Research on language sample methodology has suggested that hard and fast rules for selecting sample length are likely to cause challenges (Heilmann et al., 2010). This issue deserves special attention for children with neurodevelopmental disorders.

Despite efforts to equate certain aspects of language sampling contexts for the purposes of comparison, the adult with whom a child interacts will necessarily differ between parent-child play and examiner-child play or the ADOS. In the current study, we did not examine differences in adults' language. It might be expected that differences in adult MLU, question asking, number of utterances, etc. may correlate with aspects of child language. Such relationships have been identified for children with other neurodevelopmental disorders; yet, these effects are likely bidirectional in nature and will require future research to disentangle (Kover et al., 2012).

Here, we have examined only three language sampling contexts. Similar comparisons among the CSBS and the ESCS—other assessments suggested as potential language sampling contexts by the spoken language benchmarks framework—would also prove useful to researchers and clinicians. We also did not consider the possibility of combining multiple assessments or language sampling contexts to yield larger language samples, also suggested by the spoken language benchmarks framework. This could serve as a sensible strategy, despite the increased burden of transcription. Finally, in the current study, we did not exclude imitative language or echolalia. It may be of interest for future research to understand how echolalia varies across language sampling contexts and whether excluding these types of utterances differentially impacts classification into language phases across contexts.

## Clinical Implications

The context selected for a language sample should be determined by the goals of the assessment and may also be influenced by the total time available for assessment or concerns about a specific language domain. Taking a language sample from the ADOS or other measure of social communication reduces the time of assessment; however, a language sample drawn from examiner- or parent-child play may provide a different representation of a child's spoken language. We caution that our analyses focused on average patterns of performance rather than individual patterns; nevertheless, we did observe a tendency for more utterances and more communicative functions to be produced during parent-child play than the other contexts. Thus, interaction with a parent may facilitate an understanding of a child's spoken language under conditions of the support and scaffolding of a familiar interlocutor. Given that MLU was highest with examiner-child play, this context might be useful for tracking utterance length over time, particularly because examiner behavior can be held reasonably constant (i.e., controlled through standardization of elicitation procedures). Vocabulary ability may be maximally tapped by examiner- or parent-child play relative to other contexts. These factors are important to consider when using assessment results to determine treatment goals or to track progress, particularly because comparisons across time using different language sampling contexts are unlikely to be interpretable.

Additionally, a language sample should be coupled with information from parent report and/or direct standardized assessment in determining a child's language phase (Tager-Flusberg et al., 2009). Including multiple sources of information improves the ability to accurately assess a child's developmental profile and may ameliorate the impact of using a single language sampling context on a child's classification in the spoken language benchmarks framework. In this way, the spoken language benchmarks framework may support the selection of specific treatment goals and monitoring of developmental gains in terms of areas of relative strengths and weaknesses in relation to the research literature.

## Conclusions

Utilizing one assessment for multiple purposes (e.g., observing autism symptoms and obtaining a language sample) is an appealing way to conserve resources and minimize the size of a testing protocol in research and clinical settings; however, this benefit is only worthwhile to the extent that an assessment can serve each purpose in a satisfactory manner and in a way that is fully understood by the individual interpreting results. In the current study, we systematically compared the language performance of young children with ASD during the ADOS and examiner- and parent-child play, finding that the ADOS tended to result in lower scores for the amount of language produced, as well as structural and pragmatic language. Future studies are needed to explore how different language sampling contexts impact the conclusions that are made about change over time within the spoken language benchmarks framework.

## Acknowledgements

This research was supported by NIH grants R01 DC07223 (Susan Ellis Weismer, PI), T32 DC05359 (Susan Ellis Weismer, PI), and P30 HD03352 (Marsha Mailick, PI). The first author is now in the Department of Speech and Hearing Sciences at the University of Washington. We are grateful to the families and children who participated in

this research. Preliminary data were presented at the 2013 Symposium on Research in Child Language Disorders. We offer special thanks to Madeleine Swenson, Sarah Allen, and Jane Hohman for their tireless efforts in transcription and coding.

## References

- Barnes E, Roberts J, Long SH, Martin GE, Berni MC, Mandulak KC, Sideris J. Phonological accuracy and intelligibility in connected speech of boys with fragile X syndrome or Down syndrome. *Journal of Speech, Language, and Hearing Research*. 2009; 52(4):1048–1061.
- Bishop SL, Guthrie W, Coffing M, Lord C. Convergent validity of the Mullen Scales of Early Learning and the Differential Ability Scales in children with autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities*. 2011; 116(5):331–343. [PubMed: 21905802]
- Cohen, J. *Statistical power analysis for the behavioral sciences*. Academic Press; New York: 1969.
- Condouris K, Meyer E, Tager-Flusberg H. The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology*. 2003; 12(3):349–358. [PubMed: 12971823]
- Costanza-Smith A. The clinical utility of language samples. *Perspectives on Language Learning and Education*. 2010; 17(1):9–15.
- Dollaghan CA, Campbell TF, Tomlin R. Video narration as a language sampling context. *Journal of Speech and Hearing Disorders*. 1990; 55(3):582. [PubMed: 2381199]
- Eigsti I-M, Bennetto L, Dadlani MB. Beyond pragmatics: Morphosyntactic development in autism. *Journal of Autism and Developmental Disorders*. 2007; 37:1007–1023. [PubMed: 17089196]
- Ellis Weismer S, Gernsbacher MA, Stronach S, Karasinski C, Eernisse ER, Venker CE, Sindberg H. Lexical and grammatical skills in toddlers on the autism spectrum compared to late talking toddlers. *Journal of Autism and Developmental Disorders*. 2011; 41(8):1065–1075.
- Estigarribia B, Martin GE, Roberts JE. Cognitive, environmental, and linguistic predictors of syntax in fragile X syndrome and Down syndrome. *Journal of Speech, Language and Hearing Research*. 2012; 55(6):1600–1612.
- Estigarribia B, Roberts JE, Sideris J, Price J. Expressive morphosyntax in boys with fragile X syndrome with and without autism spectrum disorder. *International Journal of Language and Communication Disorders*. 2011; 46(2):216–230. [PubMed: 21401819]
- Evans JL, Craig HK. Language sample collection and analysis: Interview compared to freeplay assessment contexts. *Journal of Speech and Hearing Research*. 1992; 35(2):343–353. [PubMed: 1573874]
- Gardner, MF. *Expressive One Word Picture Vocabulary Test-Revised*. Western Psychological Services; Los Angeles: 1990.
- Gavin WJ, Giles L. Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech and Hearing Research*. 39:1258–1262.
- Goldman, R.; Fristoe, M. *Goldman-Fristoe Test of Articulation-2 (GFTA-2)*. American Guidance Services; Circle Pines, MN: 2000.
- Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2009; 39(5):693–705. [PubMed: 19082876]
- Haebig E, McDuffie A, Ellis Weismer S. Brief report: Parent verbal responsiveness and language development in toddlers on the autism spectrum. *Journal of Autism and Developmental Disorders*. 2013a; 43(9):2218–2227.
- Haebig E, McDuffie A, Ellis Weismer S. The contribution of two categories of parent verbal responsiveness to later language for toddlers and preschoolers on the autism spectrum. *American Journal of Speech-Language Pathology*. 2013b; 22(1):57–70.
- Hale CM, Tager-Flusberg H. Brief report: the relationship between discourse deficits and autism symptomatology. *Journal of Autism and Developmental Disorders*. 2005; 35(4):519–524.

- Hansson K, Nettelbladt U, Nilholm C. Contextual influence on the language production of children with speech/language impairment. *International Journal of Language and Communication Disorders*. 2000; 35(1):31–47. [PubMed: 10824223]
- Heilmann J, Nockerts A, Miller JF. Language sampling: Does the length of the transcript matter? *Language Speech, and Hearing Services in Schools*. 2010; 41:393–404.
- Hogan-Brown AL, Losh M, Martin GE, Mueffelmann DJ. An investigation of narrative ability in boys with autism and fragile X Syndrome. *American Journal on Intellectual and Developmental Disabilities*. 2013; 118(2):77–94. [PubMed: 23464607]
- Kover ST, McDuffie A, Abbeduto L, Brown WT. Effects of sampling context on spontaneous expressive language in males with fragile X syndrome or Down syndrome. *Journal of Speech, Language, and Hearing Research*. 2012; 55(4):1022–1038.
- Landis JR, Koch GG. Measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159–174. [PubMed: 843571]
- Le Couteur, A.; Rutter, M.; Lord, C.; DiLavore, P. *Toddler Research Autism Diagnostic Interview-Revised*. Western Psychological Services; Los Angeles: 2006.
- Lord, C.; Rutter, M.; DiLavore, P.; Risi, S. *Autism Diagnostic Observation Schedule - Generic*. Western Psychological Services; Los Angeles: 1999.
- Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Rutter M. The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*. 2000; 30(3):205–223. [PubMed: 11055457]
- Losh M, Capps L. Narrative ability in high-functioning children with autism or Asperger's syndrome. *Journal of Autism and Developmental Disorders*. 2003; 33(3):239–251.
- MacLachlan BG, Chapman RS. Communication breakdowns in normal and language learning-disabled children's conversation and narration. *Journal of Speech and Hearing Disorders*. 1988; 53:2–7. [PubMed: 3339864]
- Miller, JF.; Andriacchi, K.; Nockerts, A. *Assessing Language Production Using SALT Software: A Clinician's Guide to Language Sample Analysis*. SALT Software, LLC.; Madison, WI: 2011.
- Miller, JF.; Iglesias, A. *Systematic Analysis of Language Transcripts (SALT) (Version Research Version 2008)*. SALT Software, LLC.: 2008.
- Morley, E.; Roark, B.; van Santen, J. The utility of manual and automatic linguistic error codes for identifying neurodevelopmental disorders.. Paper presented at the Proceedings of the Eighth Workshop on Innovative Use of Natural Language Processing for Building Educational Applications; Atlanta, Georgia. 2013.
- Mullen, EM. *Mullen Scales of Early Learning: AGS Edition*. American Guidance Service; Circle Pines, MN: 1995.
- Mundy, P.; Hogan, A.; Doehring, P. *A preliminary manual for the abridged Early Social Communication Scales*. University of Miami; Coral Gables, Florida: 1996.
- Norbury CF, Bishop DV. Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders*. 2003; 38(3):287–313. [PubMed: 12851080]
- Park CJ, Yelland GW, Taffe JR, Gray KM. Morphological and syntactic skills in language samples of preschool aged children with autism: Atypical development? *International Journal of Speech-Language Pathology*. 2012; 14(2):95–108. [PubMed: 22390743]
- Paul R, Campbell D, Gilbert K, Tsiouri I. Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2013; 43(2):418–431.
- Price JR, Roberts JE, Hennon EA, Berni MC, Anderson KL, Sideris J. Syntactic complexity during conversation of boys with fragile X syndrome and Down syndrome. *Journal of Speech, Language and Hearing Research*. 2008; 51(1):3–15.
- Ray-Subramanian CE, Ellis Weismer S. Receptive and Expressive Language as Predictors of Restricted and Repetitive Behaviors in Young Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*. 2012; 42:2113–2120.

- Ray-Subramanian CE, Huai N, Ellis Weismer S. Brief report: Adaptive behavior and cognitive skills for toddlers on the autism spectrum. *Journal of Autism and Developmental Disorders*. 2011; 41(5): 679–684.
- Roberts J, Martin GE, Moskowitz L, Harris AA, Foreman J, Nelson L. Discourse skills of boys with fragile X syndrome in comparison to boys with Down syndrome. *Journal of Speech, Language, and Hearing Research*. 2007; 50(2):475–492.
- Rutter, M.; Le Couteur, A.; Lord, C. *Autism Diagnostic Interview - Revised*. Western Psychological Services; Los Angeles: 2003.
- Southwood F, Russell AF. Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech, Language, and Hearing Research*. 2004; 47(2): 366–376.
- Swensen LD, Kelley E, Fein D, Naigles LR. Processes of language acquisition in children with autism: Evidence from preferential looking. *Child Development*. 2007; 78(2):542–557. [PubMed: 17381789]
- Tager-Flusberg H, Calkins S. Does imitation facilitate the acquisition of grammar? Evidence from a study of autistic, Down's syndrome and normal children. *Journal of Child Language*. 1990; 17(3): 591–606. [PubMed: 2148571]
- Tager-Flusberg H, Calkins S, Nolin T, Baumberger T, Anderson M, Chadwick-Dias A. A longitudinal study of language acquisition in autistic and Down syndrome children. *Journal of Autism and Developmental Disorders*. 1990; 20(1):1–21.
- Tager-Flusberg H, Rogers S, Cooper J, Landa R, Lord C, Paul R, Yoder P. Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language and Hearing Research*. 2009; 52(3):643–652.
- Tager-Flusberg H, Sullivan K. Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics*. 1995; 16:241–241.
- Venker CE, Eernisse ER, Saffran JR, Ellis Weismer S. Individual differences in the real-time comprehension of children with ASD. *Autism Research*. 2013; 6(5):417–432. [PubMed: 23696214]
- Volden J, Smith IM, Szatmari P, Bryson S, Fombonne E, Mirenda P, Thompson A. Using the preschool language scale, fourth edition to characterize language in preschoolers with autism spectrum disorders. *American Journal of Speech-Language Pathology*. 2011; 20(3):200–208. [PubMed: 21478278]
- Wetherby, A.; Prizant, BM. *Communication and Symbolic Behavior Scales*. Brookes Publishing; Baltimore: 2002.
- Westerveld MF, Gillon GT, Miller JF. Spoken language samples of New Zealand children in conversation and narration. *Advances in Speech-Language Pathology*. 2004; 6:195–208.
- Zimmerman, I.; Steiner, V.; Pond, R. *Preschool Language Scale. Fourth Edition*. Psychological Corporation; San Antonio, TX: 2002.

**Table 1**

ADOS (Lord et al., 1999) Activities and Utterances Produced across Activities in 15-minute Language Samples

Activity	Modules	Mean	(SD)	Range
Free play	1, 2	22.62	(23.31)	0 - 101
Response to name	1, 2	1.94	(3.49)	0 - 19
Response to joint attention	1, 2	11.68	(13.31)	0 - 63
Bubble play	1, 2	7.65	(11.86)	0 - 47
Anticipation of a routine with objects	1, 2	3.40	(9.26)	0 - 39
Anticipation of a social routine	1	0.17	(1.07)	0 - 8
Functional and symbolic imitation	1	0.25	(1.27)	0 - 8
Birthday party	1, 2	7.71	(13.74)	0 - 58
Snack	1, 2	7.79	(21.51)	0 - 142
Construction task	2, 3	6.94	(11.48)	0 - 45
Make-believe play	2, 3	11.63	(17.00)	0 - 59
Joint interactive play	2, 3	16.65	(29.51)	0 - 123
Conversation and reporting	2, 3	2.51	(5.57)	0 - 24
Demonstration task	2, 3	2.05	(6.09)	0 - 28
Description of a picture	2, 3	4.81	(10.09)	0 - 47
Telling a story from a book	2, 3	1.62	(4.67)	0 - 20

*Note.* No analyzed utterances were produced during Social smile, Cartoons, Emotions, Social difficulties and annoyance, Break, Friends, relationships, and marriage, Loneliness, or Creating a story, which tend to occur later in ADOS administration.

**Table 2**

## Participant Characteristics

Variable	Mean	(SD)	Range
Chronological age	44.92	(3.94)	37 - 53
MSEL <sup>a</sup>			
Age-equivalent	42.23	(11.90)	24 - 69
T-score	47.27	(15.99)	20 - 79
PLS-4 Auditory Comprehension <sup>b</sup>			
Age-equivalent	32.46	(15.10)	10 - 78
Standard score	74.58	(25.58)	50 - 145
PLS-4 Expressive Communication <sup>c</sup>			
Age-equivalent	33.33	(9.33)	21 - 57
Standard score	79.12	(18.42)	50 - 125
Calibrated autism symptom severity	6.89	(1.69)	3 - 10
Mother's years of education	14.68	(2.09)	12 - 19

*Note.* Age and age-equivalent scores are given in months. MSEL = Mullen Scales of Early Learning Visual Reception subtest (Mullen, 1995). PLS-4 = Preschool Language Scale, Fourth Edition (Zimmerman et al., 2002).

<sup>a</sup>Scores were available for only 61 participants.

<sup>b</sup>Scores were available only for 59 participants.

<sup>c</sup>Scores were available for only 58 participants.

**Table 3**

Language Performance across Sampling Contexts

Variable	ADOS			Examiner-Child Play			Parent-Child Play		
	Mean	(SD)	Range	Mean	(SD)	Range	Mean	(SD)	Range
Number of utterances	109.44 <sup>a</sup>	(44.50)	26 - 208	141.44 <sup>b</sup>	(52.39)	46 - 290	179.65 <sup>c</sup>	(57.38)	62 - 315
Intelligibility	.68 <sup>a</sup>	(.15)	.19 - .98	.77 <sup>b</sup>	(.12)	.43 - .97	.78 <sup>b</sup>	(.12)	.48 - .96
NDW	62.70 <sup>a</sup>	(43.99)	5 - 186	83.76 <sup>b</sup>	(40.69)	23 - 182	82.81 <sup>b</sup>	(39.77)	22 - 201
MLU	2.10 <sup>a</sup>	(0.73)	1.00 - 4.36	2.35 <sup>b</sup>	(0.84)	1.06 - 4.73	2.11 <sup>a</sup>	(0.70)	1.14 - 3.95
Requests	14.51 <sup>a</sup>	(10.64)	0 - 55	14.41 <sup>a</sup>	(13.82)	0 - 67	18.51 <sup>b</sup>	(15.66)	0 - 60
Comments	37.21 <sup>a</sup>	(26.70)	0 - 115	57.24 <sup>b</sup>	(29.79)	0 - 128	56.49 <sup>b</sup>	(32.03)	0 - 138
Turn-taking	4.43 <sup>a</sup>	(10.03)	0 - 46	6.54 <sup>a,b</sup>	(16.74)	0 - 87	8.97 <sup>b</sup>	(19.28)	0 - 107

Note. Contexts with different superscripts differ significantly,  $p < .05$ .



**Table 4**

Frequency of Children Categorized into Each Language Phase by Domain

Context	Phonology Phase			Vocabulary Phase			Grammar Phase			Pragmatics Phase		
	1	2	3	1	2	3	1	2	3	1	2	3
ADOS	6	30	27 <sup>a</sup>	15 <sup>a</sup>	34	13 <sup>a</sup>	-	26	8	39 <sup>a</sup>	14	9
Examiner-child play	2	20	41	2 <sup>b</sup>	35	26 <sup>b</sup>	-	36	11	33	21	8
Parent-child play	2	21	40 <sup>b</sup>	2 <sup>b</sup>	36	25 <sup>b</sup>	-	31	8	27 <sup>b</sup>	22	12

*Note.* Phase 1 = First Words; Phase 2 = Word Combinations; Phase 3 = Sentences. A dash indicates that the First Words language phase does not apply to the Grammar domain. One participant was below the level of First Words for Vocabulary for the ADOS, Pragmatics for the ADOS, and Pragmatics for examiner-child play. Two participants were below the level of First Words for Pragmatics for parent-child play. For Grammar, 29 participants during the ADOS, 16 participants during examiner-child play, and 24 participants during parent-child play failed to reach Word Combinations. Within columns, contexts with different superscripts differ significantly in the number of participants categorized into that phase versus others.