Title

Narrative Language Sampling in Typical Development: Implications for Clinical Trials

Journal

American Journal of Speech-Language Pathology, 27(1)

ISSN

1058-0360

Authors

Channell, Marie Moore
Loveall, Susan J
Conners, Frances A
et al.

Publication Date

2018-02-06

DOI

10.1044/2017_ajslp-17-0046

Peer reviewed

AJSLP

Research Article

Narrative Language Sampling in Typical Development: Implications for Clinical Trials

Marie Moore Channell,[a] Susan J. Loveall,[b] Frances A. Conners,[c] Danielle J. Harvey,[d] and Leonard Abbeduto[d]

Purpose: This study examined cross-sectional age-related trajectories of expressive language variables (syntactic complexity, lexical diversity, unintelligibility, dysfluency, and talkativeness) derived from a narrative language sampling procedure.
Method: Narrative samples were analyzed from 103 typically developing individuals, ages 4–21 years.
Results: Results showed that this procedure was effective for the entire age range, with participants producing an utterance on virtually every page of the wordless picture books used to prompt the narrative. Importantly, the cross-sectional trajectories for syntactic complexity and lexical diversity showed age-related increases through the age of 18 years, although measures of other dimensions of language showed different relationships with age.
Conclusions: These data inform developmental work and document the extent to which the narrative procedure can be used to characterize expressive language over a wide age range. This procedure has been proposed as an outcome measure for clinical trials and interventions involving individuals with intellectual and developmental disabilities. The present data document the developmental levels for which the procedure and metrics derived are appropriate.

Individuals with intellectual and developmental disabilities of various etiologies (e.g., Down syndrome, fragile X syndrome, autism spectrum disorder) experience a range of difficulties with various aspects of communication that almost invariably include spoken language impairments (Abbeduto, McDuffie, Thurman, & Kover, 2016; McDuffie, Thurman, Channell, & Abbeduto, 2017). Fortunately, in recent years more treatment options have become available for these individuals, including behavioral interventions that target specific aspects of language and pharmaceutical therapies that target common issues associated with different disorders (e.g., social anxiety, hyperactivity, inattention; Hagerman & Hendren, 2014). Although not always a primary end point in studies of the efficacy of these pharmaceutical agents, improvements in spoken language have often been noted in clinician or parent report (Berry-Kravis, Doll, et al., 2013). As new treatments are being offered to individuals with intellectual and developmental disabilities, there is a growing need for outcome measures that are relatively quick and easy to administer, are sensitive to change, and capture clinically meaningful outcomes, including those in the domain of spoken language (Berry-Kravis, Hessl, et al., 2013; Esbensen et al., 2017). This is particularly important, given the extent of the spoken language impairments that impact everyday communication and social interaction in individuals with intellectual and developmental disabilities (Abbeduto et al., 2016).

The search for outcome measures, however, has been slowed by a lack of appropriate data regarding, among other things, the typical developmental trajectory of the measure, which is essential for understanding potential floor and ceiling effects and the ability range for which the measure is appropriate. This study provides such data for a narrative language sampling procedure developed by Abbeduto, Benson, Short, and Dolish (1995). This procedure has been proposed as a potential outcome measure for clinical trials involving individuals with intellectual and developmental disabilities (Berry-Kravis, Hessl, et al.,

[a]University of Illinois at Urbana-Champaign
[b]University of Mississippi, Oxford
[c]University of Alabama, Tuscaloosa
[d]University of California, Davis, MIND Institute, Sacramento
Correspondence to Marie Moore Channell: channell@illinois.edu

Editor-in-Chief: Krista Wilkinson
Editor: Carol Miller
Received April 12, 2017
Revision received June 29, 2017
Accepted September 5, 2017
https://doi.org/10.1044/2017_AJSLP-17-0046

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

2013; Budimirovic et al., 2017; Esbensen et al., 2017) and has been used in several clinical trials for these populations (e.g., Protocol Neu-2566-FXS-001; *A Safety Study of NNZ-2566 in Patients With Fragile X Syndrome, 2013–2017*).

Limitations of Standardized Norm-Referenced Measures

Although standardized norm-referenced assessments of cognition and language are valuable tools for many populations, they present particular challenges when used in populations with intellectual disabilities. One of the greatest challenges is that many individuals with intellectual disabilities perform at floor level in terms of the standard scores they achieve because such tests frequently are not normed much beyond 2 *SD*s below the mean (Hessl et al., 2009). This is problematic not only because it can lead to inadequate conclusions regarding true ability levels, but also because it impedes the assessment of change over time. That is, an individual who achieves a standard score at the floor at the time of the first assessment and again at the second assessment may in fact have progressed in skills between the two time points, but the assessment fails to capture this important change, thereby making the measure inadequate for many treatment studies. Although newer editions of standardized measures are beginning to include growth scores that can be used to track progress over time (i.e., gain in raw abilities) without vulnerability to floor effects, they do not provide information relative to same-age peers, making them qualitatively different from standard scores. Furthermore, growth scores are not yet widely available for many of the most frequently used standardized language measures; thus, the limitations in standardized norm-referenced measures persist.

Another limitation of many standardized norm-referenced assessments of language, in particular, is that individuals with intellectual disabilities often have unique patterns of impairment across different domains of language, making the composite scores that concatenate across multiple domains that are provided by many standardized language assessments difficult to interpret. For instance, individuals with Down syndrome tend to struggle particularly with expressive syntax relative to expressive vocabulary (Chapman & Hesketh, 2000), whereas those with Williams syndrome display expressive syntax in advance of some aspects of vocabulary, such as relational words (Mervis & John, 2010). A single composite score of expressive language, then, would not capture such differences and thus would be of limited utility in a treatment study. Finally, standardized language assessments rely on test formats that are unlike real-world language use (e.g., completing a sentence fragment or generating a response to a hypothetical social situation), and thus, their generalizability to meaningful language use is also often limited.

Utility of Expressive Language Sampling Procedures

Expressive language sampling procedures circumvent many of the challenges facing standardized tests for assessing expressive language ability and hold considerable promise for use as an outcome measure across many populations with communication disorders, including those with intellectual disabilities (Abbeduto, Kover, & McDuffie, 2012). In fact, expressive language sampling procedures have been shown to be effective at identifying children with language impairments, even in the absence of low IQs (e.g., as in specific language impairment; Heilmann, Miller, & Nockerts, 2010; Rescorla, Roberts, & Dahlsgaard, 1997; Rice, Redmond, & Hoffman, 2006; Rice et al., 2010). Furthermore, if anything, these procedures are less subject to identification bias in ethnic and racial minority groups than traditional standardized norm-referenced language assessments (Craig & Washington, 2000; Heilmann & Westerveld, 2013; Mills, Mahurin-Smith, & Steele, 2017).

In such expressive language sampling procedures, the idea is to collect a sample of spoken language in a naturalistic context in such a way that the sample is representative of the individual's "everyday" language abilities. Representativeness can be influenced by a number of contextual factors, including the genre, topic, and behavior of conversational partners (Abbeduto et al., 1995; Johnston, Miller, Tallal, & Curtiss, 1993; Leadholm & Miller, 1992; Morris-Friehe & Sanger, 1992; Wagner, Nettelbladt, Sahlén, & Nilholm, 2000). From the perspective of outcome measure development, it is essential that these factors are kept as constant as possible across participants and occasions of assessment (Berry-Kravis, Doll, et al., 2013; Berry-Kravis, Hessl, et al., 2013). It is particularly important to script the behavior of the conversational partner to ensure that it does not decrease the representativeness of the sample with regard to the participant's language abilities (Abbeduto et al., 1995). For example, excessive use of yes/no questions may lead the participant to produce shorter and less complex utterances than he or she would otherwise, whereas excess prompting and scaffolding may lead to output that appears more sophisticated than is typical for the participant in the less scaffolded interactions of everyday life. At the same time, however, it is essential to balance the need for such "standardization" of the sampling context with the need to keep the interaction natural and thereby ensure that performance is generalizable to meaningful everyday communicative interactions (Abbeduto et al., 2012).

Methods for creating consistency in administration have been developed for several expressive language sampling procedures and for several populations with, or at risk for, language disorders (Heilmann et al., 2008), including those with accompanying intellectual disabilities (Abbeduto et al., 2012). Because of the relative ease of administration (Berry-Kravis, Doll, et al., 2013), language-sampling procedures in the context of narration appear especially promising as outcome measures for treatment studies involving individuals with intellectual disabilities (Berry-Kravis, Hessl, et al., 2013; Budimirovic et al., 2017; Esbensen et al., 2017). Narrative procedures involving wordless picture book stimuli, specifically, are relatively easy to standardize in terms of examiner behavior and content due to the structured nature of the picture books (Abbeduto et al., 2012). The context of narration is also particularly appealing for use in populations with intellectual disabilities, because it can be used in wide age ranges and ability levels. For example, wordless picture books can be selected that are appropriate for a wide range of ages, avoiding the problem of materials that are too abstract for young children and too "babyish" for adolescents and adults. Moreover, storytelling is an activity that occurs at all ages from the preschool years through adulthood, making it ecologically valid across virtually the entire life span. Furthermore, in narration (i.e., coherently relaying causally and/or temporally related past events), the same narrative events can be described in concrete or abstract ways, with simple or complex syntax, and so forth (Abbeduto et al., 2012), again speaking to its applicability to wide age ranges.

Narrative procedures have yielded replicable findings documenting cross-linguistic developmental differences and similarities (Berman & Slobin, 1994) as well as disorder-specific profiles of language impairment for several common conditions associated with intellectual disabilities (Abbeduto et al., 2016). Because narration serves as a showcase for multiple dimensions of language, once transcribed, the expressive language samples can be analyzed for a wide array of language skills, including microstructural (e.g., grammatical word categories, mental state language) and macrostructural (e.g., story grammar, evaluations) domains. The ability to derive scores for multiple language domains is key, given the variability of individuals with intellectual disabilities of known etiology (e.g., Down syndrome, fragile X syndrome, Williams syndrome) who often demonstrate unique profiles across these domains (Abbeduto et al., 2016). As a first step, the purpose of the current study focuses specifically on certain microstructural metrics derived from narrative language samples that can be computed automatically (as opposed to those requiring hand coding), maximizing their appeal for use in clinical trials (Esbensen et al., 2017).

Eliciting Narratives

We employed the procedure developed by Abbeduto and colleagues (e.g., Abbeduto et al., 1995; Berry-Kravis, Doll, et al., 2013) from procedures commonly employed to collect narrative language samples from speakers across numerous cultures and languages using wordless picture books (e.g., Berman & Slobin, 1994; Ganthous, Rossi, & Giacheti, 2017; Heilmann, Rojas, Iglesias, & Miller, 2016; Hoffman, 2009). Like most variants of these procedures, the procedure developed by Abbeduto and colleagues minimizes examiner scaffolding. However, Abbeduto and colleagues' adaptation differs from other variants of the procedure largely by a greater specification of the examiner's behavior with regard to the frequency, types, and sequence of prompts, as well as in the timing of the pacing of the procedure. This increased structure was largely motivated by the goal of standardization of use in populations who have intellectual and developmental disabilities, including use in multisite clinical trials. In fact, this adapted procedure has been suggested as a potentially useful outcome measure for treatment studies in several atypically developing populations (Budimirovic et al., 2017; Esbensen et al., 2017). The Abbeduto et al. procedure also has been used successfully to describe the language impairment profiles of individuals with several conditions associated with intellectual and developmental disabilities, including Down syndrome, fragile X syndrome, and autism spectrum disorder (Abbeduto et al., 1995; Berry-Kravis, Doll, et al., 2013; Finestack & Abbeduto, 2010; Keller-Bell & Abbeduto, 2007; Kover, McDuffie, Abbeduto, & Brown, 2012). Preliminary findings also suggest that scores derived from this procedure and reflecting several important dimensions of language (e.g., syntax; vocabulary) have strong psychometric properties from the perspective of treatment studies, including excellent test–retest reliability for individuals with intellectual disabilities (Berry-Kravis, Doll, et al., 2013). The Abbeduto et al. narrative procedure is also being used in several ongoing clinical trials involving these diagnostic conditions.

No study to date, however, has systematically examined the age-related trajectories of outcome variables derived from the Abbeduto et al. narrative procedure in typically developing individuals. Consequently, the ages and developmental levels at which the procedure elicits variability in aspects of expressive language, such as syntax and vocabulary, is not known. These data are critical for deciding not only the appropriateness of the procedure itself for a clinical trial but also the outcome variables to be derived from it to assess treatment efficacy. For example, does the variable to be derived have the potential to show change for the developmental range of the participants? Importantly, simply aggregating data across participants producing narratives based on different stories, materials, or elicitation procedures would not provide the sort of data needed to guide the use of any given narrative task in clinical trials. Accordingly, this study was designed to provide the first data on a large sample of typically developing individuals across a wide age range (4–21 years) by examining a set of expressive language variables derived from the Abbeduto et al. narrative procedure that index syntactic complexity (i.e., mean length of utterance [MLU]), lexical diversity, unintelligibility, dysfluency, and talkativeness (i.e., number of utterances attempted per minute).

Expressive Language Variables Derived From the Abbeduto et al. Narrative Procedure

Five variables were selected to index different features of expressive language, some of which are expected to reflect developmental changes with age (at least throughout childhood) and others that distinguish typical children from those with various types of language disorders. As Miller (1991) first proposed, these variables can be considered metrics of "developmental progress" and/or "disordered performance." These outcome variables also are particularly attractive for use in clinical trials because of their collective breadth and the fact that they are computed automatically from the most commonly used software packages for transcription (e.g., Systematic Analysis of Language Transcripts [SALT]; Miller & Iglesias, 2006), which increases feasibility in large multisite trials (Berry-Kravis, Doll, et al., 2013; Budimirovic et al., 2017; Esbensen et al., 2017).

In terms of validity of the metrics of "developmental progress," Miller and colleagues have reported findings from expressive language samples of several cohorts of children ages 3–13 years (currently available in the SALT reference database) that psychometrically support the use of MLU as a metric of syntax and lexical diversity as a metric of semantics and that document developmental changes in these metrics across the age range, particularly in the context of narration (Heilmann et al., 2010; Leadholm & Miller, 1992; Miller, 1991; Miller et al., 2005; Miller & Klee, 1995). Other research groups have documented similar findings in different sampling contexts (e.g., Rice et al., 2006, 2010), supporting the stability of these findings across contexts (e.g., conversation, narrative, expository discourse), and have emphasized the need for creating age-related developmental trajectories for such variables of interest (Rice et al., 2006). No study to date, however, has reported age-related trajectories in these variables in the context of narration across a wide range, from early childhood into adulthood. The current study addresses this aim and does so for ages 4–21 years, allowing for a thorough examination of variability and documentation of potential floor and ceiling effects.

Examining the validity of the metrics that distinguish typical from language-disordered performance, prior work has again demonstrated MLU and lexical diversity as good indicators (Heilmann et al., 2010; Miller & Klee, 1995; Scott & Windsor, 2000). In addition, fluency and rate of talk have surfaced as other indicators of "disordered performance." Specifically in the context of narration, number of mazes, pauses, and other aspects of dysfluency along with speaking rate (e.g., number of words per minute) have distinguished typical children from those with various types of language impairment (Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004; Heilmann et al., 2010; Scott & Windsor, 2000). In populations with intellectual disabilities, who by definition experience cognitive impairments, both dysfluency—as a general indicator of language-planning problems—and rate of talk can be expected to be impacted and are important aspects of discourse to consider. Furthermore, although not a challenge for typically developing children by school entry, limited intelligibility is a serious challenge for many individuals with intellectual disabilities, from Down syndrome to fragile X syndrome, because of their poor articulation and phonological problems (e.g., Kover et al., 2012). Thus, in addition to MLU and lexical diversity, dysfluency, rate of talk (i.e., "talkativeness"), and unintelligibility are important variables to include, given the goal of informing clinical trials in our current study.

Current Study

The primary aim of this study was to examine cross-sectional age-related developmental trajectories for each of the five aspects of spoken language indexed by the Abbeduto et al. narrative procedure. These data will provide key information regarding the versatility of this adapted procedure by detailing the extent to which it captures abilities in each aspect of spoken language across the different ages and, thus, ability levels represented in our sample. We expect to observe improvement in each metric with age, although the rate and timing of each trajectory may vary; examining the trajectories separately for each aspect of language allows for such expected variation.

These data will set potential floors and ceilings of the variables in ways that can be used in clinical trials and treatment studies to match measures to the sample of participants' levels of functioning. For example, for an individual entering a clinical trial, it will be helpful to know the extent to which this narrative procedure is appropriate for documenting different aspects of expressive language (e.g., MLU, lexical diversity), given the individual's age and developmental level. In addition, the inclusion of a wide age range of typically developing individuals will contribute to our understanding of the microstructural language components of narrative discourse from a developmental perspective.

A secondary aim was to investigate potential sex differences in these cross-sectional age-related trajectories. Although there is limited support in the literature for broad sex differences in language abilities beyond toddlerhood (Bornstein, Hahn, & Haynes, 2004), an important question remains whether there are sex differences in the age-related increases across the five indexed aspects of spoken language. In addition, there is some evidence of sex differences in performance across language domains and underlying neurological structures in certain disorders that carry unequal prevalence rates between sex (e.g., autism spectrum disorder, Nordahl et al., 2015; fragile X syndrome, Abbeduto et al., 2003; Murphy & Abbeduto, 2007; Pierpont, Richmond, Abbeduto, Kover, & Brown, 2011) as well as in intellectual disabilities more generally (Maulik, Mascarenhas, Mathers, Dua, & Saxena, 2011). Because these disorders are commonly targeted in treatment studies that may adopt the Abbeduto et al. narrative procedure as an outcome measure, it is important to document any potential sex differences in the variables of interest across the age range in typically developing individuals as well.

Method

Participants

Participants were drawn from a larger study on cognitive predictors of language impairment in youth and young adults with intellectual disabilities. The larger study included samples of typically developing individuals who served as control participants for the groups with intellectual disabilities. The current study examined available data from all typically developing individuals who participated in the larger study.

Criteria for inclusion in the typically developing group of the larger study were as follows: 4–21 years of age, not currently or previously receiving special education services (including those for learning disability, speech/language therapy, and/or giftedness), without a parent-reported diagnosis of autism spectrum disorder or attention-deficit hyperactivity disorder, and a native English speaker. Recruitment and assessment of typically developing participants were conducted at the lead university for the project, which is located in the southeastern United States. Typically developing participants were recruited via local preschools (including the university preschool center), the local public school district, and the university psychology undergraduate subject pool. A total of 112 individuals met these criteria and were enrolled in the larger study.

In addition, to be included in the present report, participants were required to have completed the Abbeduto et al. narrative procedure. Two participants withdrew from the study without providing a reason prior to completing the narrative procedure, and one participant began the narrative procedure but refused to complete it. For an additional five participants, the audio recorder malfunctioned, resulting in missing data. Finally, to further ensure that our sample was composed of individuals with typical development status, we excluded any participant who scored greater than 1 *SD* below the mean on two or more of the three standardized measures of receptive language and nonverbal cognition, which resulted in the exclusion of one additional participant. To avoid being overly restrictive, however, we included any participant who scored beyond 1 *SD* below the mean on only one of the three standardized measures.

Our final sample consisted of 103 typically developing children, adolescents, and young adults. The sample was 62% female and 38% male. They were 62% White Non-Hispanic, 22% Black Non-Hispanic, 5% White-Hispanic, 5% Black-Hispanic, 1% Asian/Pacific Islander, 4% multiracial, and 1% unreported. The highest levels of maternal education achieved for the sample were 1% completing less than eighth grade, 1% completing eighth grade, 14% completing high school, 23% completing some college, 28% graduating from college, 13% completing some graduate-level/professional coursework, and 20% graduating with a graduate-level or professional degree. Overall, our sample consisted of a slightly higher percentage of mothers with college or graduate degrees than in the U.S. population reported by the U.S. Census Bureau in 2015, which was 19% and 11%, respectively (U.S. Census Bureau, 2015). See Table 1 for additional descriptive characteristics.

Procedure

In the larger study, participants were administered a more extensive battery of tests across two to four sessions, depending on the needs of the individual participant. The Leiter International Performance Scale–Revised (Leiter-R; Roid & Miller, 1997) and the Test of Reception of Grammar–Second Edition (TROG-2; Bishop, 2003) were always administered in the first session, and the Abbeduto et al. narrative procedure and the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007) were part of the battery administered in a subsequent session.

Abbeduto et al. Narrative Procedure

This adapted narrative language-sampling procedure, developed by Abbeduto and colleagues (e.g., Abbeduto et al., 1995; Berry-Kravis, Doll, et al., 2013; Finestack & Abbeduto, 2010; Kover et al., 2012), was administered to participants as a tool for assessing their expressive language abilities in a naturalistic context (i.e., narration) that is also standardized across participants. In this narrative procedure, participants were shown either of two wordless picture books by Mercer Mayer—*Frog Goes to Dinner* (Mayer, 1974) or *Frog on His Own* (Mayer, 1973)—with book selection counterbalanced across participants (51% received *Frog Goes to Dinner,* 49% received *Frog on His Own*). For the initial viewing of the book, each participant was instructed to look at the pictures on each page to find out what happens in the story, but without talking; the examiner controlled the length of exposure by turning the page every 10–12 s. Then, the examiner instructed the participant to tell the story to her while viewing the book a second time. The examiner waited 5–7 s after the participant finished talking about a page before turning to the next page. This delay was determined to be the appropriate balance between allowing for additional processing time to generate talk about the story while still remaining responsive to the participant by moving forward to the next page after the participant stopped talking. Examiners used the same standardized hierarchy of prompts for all participants, as needed, which minimized any scaffolding of the retell after the first page (see Appendix for prompting scripts). Participants' narratives were digitally audio-recorded for later transcription.

Transcription

Trained personnel transcribed the audio files of participants' language samples using the SALT (Miller & Iglesias, 2006) software. For each language sample, a primary transcriber completed the initial transcription, and a second transcriber checked the first draft and provided feedback for the primary transcriber, who completed the final transcript. Participants' speech was segmented into communication units (C-units); a C-unit is an independent clause with its modifiers, including dependent clauses (Loban, 1976). Segmentation into C-units yields a more accurate measure of language ability for children whose expressive language is beyond that of an average 3-year-old (Abbeduto et al., 1995). Transcribers were highly trained according to the procedures described by Abbeduto and colleagues (1995).

Table 1. Participant descriptive statistics.

Characteristic	Mean	(*SD*)	Range
Age	12.06	(5.34)	4.17–20.83
Leiter-R Nonverbal Brief IQ Composite	101.85	(13.45)	76–135
TROG-2 Standard Score (Receptive Grammar)	103.10	(11.52)	65–145
PPVT-4 Standard Score (Receptive Vocabulary)	105.92	(11.81)	75–140

Note. Leiter-R = Leiter International Performance Scale–Revised; TROG-2 = Test of Reception of Grammar–Second Edition; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition.

Ten of the language samples from the participants included in this study were randomly selected for independent transcription to ensure lack of drift in transcription procedures across the life of the project. Only dimensions of transcription relevant to the dependent measures of this study were assessed. Intertranscriber agreement averaged 86% for segmentation into C-units, 99% for identification of unintelligible C-units, 99% for identification of complete C-units, 96% for identification of C-units containing mazes, 81% for identification of the exact number of morphemes in each C-unit, 87% for identification of the exact number of words in each C-unit, and 87% for the exact lexical and morphemic content of each C-unit. Note that the last three dimensions required that the two transcriptions were in complete agreement for a C-unit, which was a conservative approach to computing agreement. For example, a C-unit represented as nine morphemes in one transcript and as 10 in the second transcript would simply be scored as a disagreement (i.e., 0% agreement rather than as 90% agreement), and a C-unit transcribed as "*the boy chased a car*" in one transcript and as "*the boy chased the car*" in the second transcript would be scored as a disagreement (i.e., 0% agreement rather than as 80% agreement). More detailed procedures regarding the rationale for and computation of agreement for these dimensions of transcription are available from the authors by request.

Primary Expressive Language Variables

SALT software was used to compute five outcome metrics of expressive language derived from transcripts of participants' narrative language samples. Each variable is described below.

Talkativeness. The number of C-units attempted per minute was used to measure the participant's rate of talk during the language sample. Higher scores indicate more talkativeness.

Unintelligibility. The proportion of C-units containing one or more segments that were unintelligible (to the transcribers) was calculated as a measure of the participant's speech intelligibility, a reflection of articulation problems. Higher scores indicate less intelligible speech (i.e., greater articulation difficulties).

Syntactic Complexity. Each participant's mean length of C-unit (MLU) in morphemes was used as a measure of complexity of expressive syntax. We included onlycomplete and fully intelligible C-units in this calculation because it is not possible to determine how many morphemes were produced in a C-unit that includes one or more unintelligible segments. Higher scores reflect greater syntactic complexity of the participant's expressive language.

Lexical diversity. The number of different word roots produced in the participant's first 50 complete and fully intelligible C-units was calculated. We used only the first 50 C-units produced by each participant to control for variability in length of participants' narratives. If a participant produced fewer than 50 C-units in his or her language sample, the total number of different word roots produced was taken as the participant's score. Sixty-five participants produced narratives composed of fewer than 50 C-units. Higher scores reflect more lexical diversity of the language sample, indicating greater expressive vocabulary.

Dysfluency. The percentage of C-units containing a maze or verbal dysfluency (e.g., filled pauses, false starts, abandoned utterances) was used to measure the dysfluency of the participant's language, a reflection of difficulty in language planning. Higher scores indicate more dysfluent speech.

Other Measures

Nonverbal Cognition

The Brief IQ subtests of the Leiter-R (Roid & Miller, 1997) were administered as a standardized measure of nonverbal cognition. The Leiter-R is nonverbal in administration and in participant response method. The Leiter-R Brief IQ screener correlates with the Wechsler Intelligence Scale for Children–Third Edition (Wechsler, 1991) at $r = .85$, and the reported test–retest reliability is $r = .88$. The Leiter-R was normed for ages 2–21 years. We used Brief IQ standard scores to describe our participant sample.

Receptive Syntax

The TROG-2 (Bishop, 2003) is a standardized measure of receptive syntax. Participants were instructed to point to pictures that best represented the phrases or sentences spoken by the examiner. The TROG-2 is norm-referenced for ages 4–86 years, and the reported internal consistency reliability is .88. We used standard scores to describe our participant sample.

Receptive Vocabulary

The PPVT-4 (Dunn & Dunn, 2007) is a standardized test of receptive vocabulary. Participants were instructed to point to pictures that best represented each word spoken by the examiner. The PPVT-4 is normed for ages 2.5–90 years and older, with reported internal consistency of .94 and test–retest reliability of .93. We used standard scores to describe our participant sample.

Results

Preliminary Analyses

The majority (i.e., 75%) of our sample produced narratives that included no unintelligible or partly unintelligible C-units; thus, unintelligibility was not analyzed further. Based on visual inspection of the plots, one participant who was an outlier for syntactic complexity was determined to have undue influence over the pattern of results. The outlier was removed in all analyses that included this variable.

Data Analysis Plan

To address our primary aim, we used linear regression to estimate the relationship between age and each of the four remaining dependent variables derived from the narrative language samples. See Table 2 for the descriptive statistics of these variables in our sample.

To address our secondary aim of examining sex differences, for each of the models representing the relationship between age and one of the four expressive language variables, we examined whether there were any sex differences. Because there was, unexpectedly, a significant difference in the ages represented by male participants (M = 10.02, SD = 5.22) and female participants (M = 13.31, SD = 5.05) in our sample, $t(101) = 3.16$, $p = .002$, we examined (a) whether sex explained a significant amount of variance beyond that attributed to age and (b) whether there was a significant interaction between sex and age for each of the variables.

Table 2. Expressive language variables.

Variable	Mean	(*SD*)	Range
Syntactic complexity (MLU)	9.55	(2.18)	3.49–14.09
Lexical diversity	113.21	(27.06)	47–183
Dysfluency	26.42	(14.76)	0.00–86.00
Talkativeness	9.06	(2.39)	4.33–20.52
Total number of C-units	47.10	(20.76)	24–165
Total time elapsed (min)	5.43	(1.79)	3.00–13.27

Note. n = 102 for MLU due to outlier removed; n = 103 for all other measures. MLU = mean length of utterance; C-units = communication units.

Age and Expressive Language Variables

Talkativeness

There was not a significant relationship between age and talkativeness, $F(1, 101) = 0.11$, $R^2 = .001$, $p = .74$.

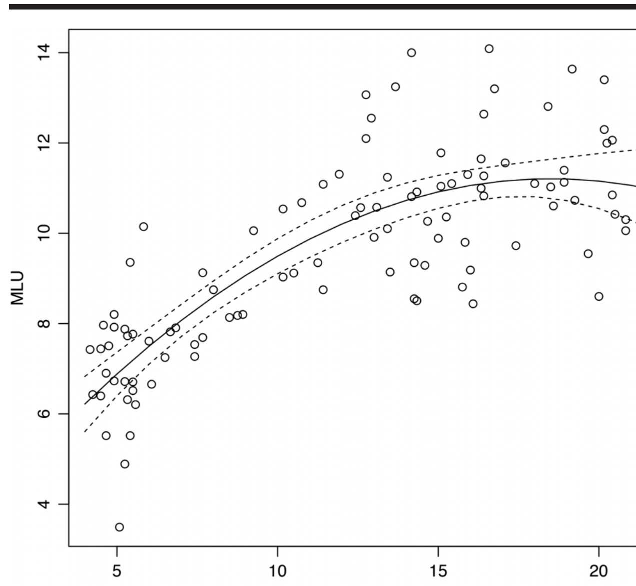
Syntactic Complexity

The relationship between age and MLU violated the assumption of linearity; thus, once the linear regression equation was estimated, a quadratic function of age was also obtained by adding an age-squared term to address the nonlinearity. The model including both age and age-squared met the assumptions of linear regression and was statistically significant, $F(2, 99) = 88.04$, $p < .001$, $R^2 = .64$ (age $\beta = 1.57$, $p < .001$; age-squared $\beta = -0.85$, $p < .001$), accounting for 64% of the total variance in MLU and indicating a quadratic trajectory of syntactic complexity with age (Figure 1). Next, we calculated the maximum of the estimated function in the model, which revealed that there was a plateau at the age of 18.5 years (i.e., no additional increase in MLU with increase in age).

Lexical Diversity

The relationship between age and lexical diversity also was nonlinear; again, a quadratic function of age was estimated. The model including age and age-squared met the assumptions of linear regression and was statistically significant, $F(2, 100) = 19.38$, $p < .001$, $R^2 = .28$ (age $\beta = 1.04$, $p < .001$; age-squared $\beta = -0.56$, $p = .05$), accounting for 28% of the total variance in lexical diversity and indicating a quadratic trajectory with age (Figure 2). Next, we calculated the maximum of the estimated function in the model, which revealed a plateau at the age of 18.5 years.

Figure 1. Cross-sectional trajectory of syntactic complexity (mean length of utterance [MLU]) across age. The dashed lines represent 95% confidence bands.



Note that because 63% of our sample produced fewer than 50 C-units, we also conducted the regression analyses for lexical diversity, adding the total number of complete and fully intelligible C-units as a covariate, and the pattern of results did not change.

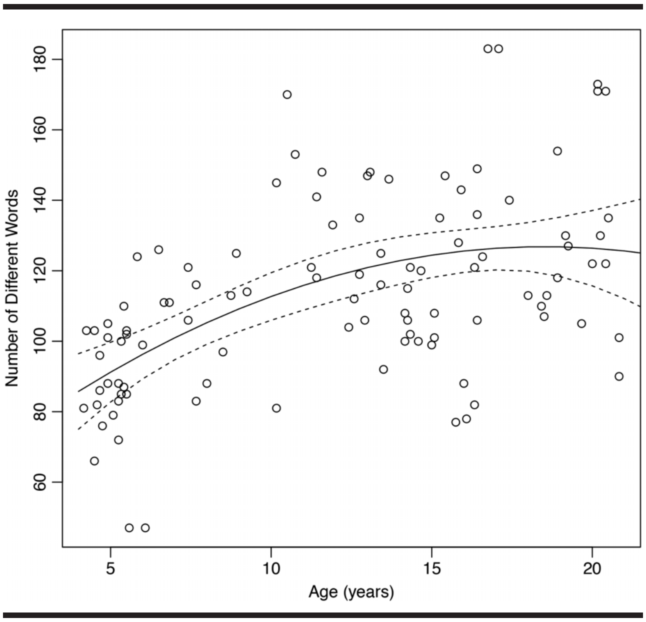
Dysfluency

The relationship between age and dysfluency was linear, positive, and statistically significant, $F(1,101) = 4.75$, $p = .03$, $R^2 = .05$, indicating a linear decrease in dysfluency across the full age range. Age, however, accounted for only 5% of the total variance in dysfluency.

Sex Differences

There was no significant main effect of sex in any of the models: talkativeness ($\beta = -.09$, $t = -0.83$, $p = .41$), MLU ($\beta = -.06$, $t = -0.95$, $p = .35$), lexical diversity ($\beta = .01$, $t = 0.08$, $p = .94$), or dysfluency ($\beta = .09$, $t = 0.83$, $p = .41$). Likewise, the interaction terms were not significant for talkativeness (Sex × Age, $p = .96$), MLU (Sex × Age, $p = .21$; Sex × Age-Squared, $p = .08$), or dysfluency (Sex × Age, $p = .91$). For lexical diversity, however, there was a marginally significant interaction between sex and age ($p = .05$), which reflected a slight difference in the variation of performance between male and female participants across the ages represented in the trajectory. The interaction between sex and age-squared, however, was not significant ($p = .12$), and the main effects of age and age-squared in the model, including both interaction terms, were also not significant ($p = .27$ and .81, respectively). Thus, there was not a consistent difference between sex in lexical diversity across the age range. Furthermore, after adding the total number of complete and fully intelligible C-units as a covariate to the model of lexical diversity, the Sex × Age interaction was no longer marginally significant.

Figure 2. Cross-sectional trajectory of lexical diversity (number of different words) across age. The dashed lines represent 95% confidence bands.



Discussion

The purpose of this study was to examine cross-sectional age-related trajectories in each of five aspects of spoken language indexed by the Abbeduto et al. narrative procedure—syntactic complexity (MLU), lexical diversity, unintelligibility, dysfluency, and talkativeness—in a large sample of typically developing children spanning 4–21 years of age. Determining how this adapted narrative procedure characterizes the emergence and cross-sectional age-related change of these abilities in typical development is crucial to understanding its utility in populations with communication disorders. More specifically, it will allow for appropriate matching of these expressive language variables to the developmental levels of the samples included in clinical treatment studies gauging treatment efficacy.

We examined the relationship between age and each aspect of spoken language separately to allow for varying cross-sectional trajectories. First, addressing the variables that differentiate typical from “disordered performance,” we examined the relation between age and unintelligibility, dysfluency, and talkativeness. As would be expected for a typically developing sample, approximately 75% of the narratives produced by our participants were fully intelligible, so we did not examine the trajectory of the unintelligibility variable further. However, the 25% of narratives that included unintelligible C-units were produced largely by the youngest participants in the sample, with the median age of the participants with some unintelligible C-units being 7.3 years and that of the fully intelligible participants being 13.6 years, fitting with developmental expectations of speech intelligibility. Notably, unintelligibility may still be a useful variable to include when applying the Abbeduto et al. narrative procedure to samples with intellectual disabilities, even at older ages, at least for those disorders in which we expect unintelligibility to be a problem (e.g., Down syndrome) because it plays an important role in effective communication.

We observed a statistically significant relationship between age and dysfluency, such that dysfluency decreased with age, consistent with expectations in typical development. Age, however, accounted for only 5% of the total variance in dysfluency, likely due to the low levels of dysfluency characteristic of the majority of ages in our sample; this is consistent with prior work (Leadholm & Miller, 1992). Again, this is an important variable to consider when examining discourse in individuals with developmental disorders of all ages, given that dysfluency is common, particularly in complex discourse contexts such as narration.

Interestingly, there was not a significant relationship between age and talkativeness derived from the Abbeduto et al. narrative procedure. However, talkativeness was calculated using the total number of utterances attempted, and it is likely that, in this typically developing sample, as the older children and adolescents used more complex utterances, they were able to effectively communicate their narratives using the same number of utterances as the younger children. Future work could consider including a different metric, such as clausal elaboration, although it would require hand coding (e.g., the subordination index) and thus could be more difficult to apply to clinical trials. Nevertheless, rate of talk is another important aspect of language that is impacted by language disorders and distinguishes typical from disordered performance (see Fey et al., 2004; Heilmann et al., 2010; Scott & Windsor, 2000); thus, talkativeness is still important to consider when using this adapted narrative procedure in clinical populations.

Next, to address the metrics of "developmental progress" in addition to "disordered performance," we examined the cross-sectional age-related trajectories of syntactic complexity (i.e., MLU) and lexical diversity (i.e., number of different words). Not surprisingly, both syntactic complexity and lexical diversity were significantly related to age. Interestingly, this relationship was nonlinear and best represented by a quadratic model for both variables, so we further examined each model to determine the maximum value of the curve—that is, at what ages the increase in each aspect of language plateaued. This plateau occurred at 18.5 years for both syntactic complexity and lexical diversity, suggesting that the narrative procedures we used showcased continually developing syntactic and semantic skills through adolescence. These results support work by Nippold and colleagues that have demonstrated such continued increase into adolescence (Nippold, 2010), primarily in the context of expository discourse (Nippold, Hesketh, Duthie, & Mansfield, 2005) and in narrative discourse using more complex stories such as fables (Nippold et al., 2014). Our results extend this work to the context of a simple story generation that can be used both in young children and adolescents as well as in individuals with developmental ages in these ranges. Furthermore, our results complement recent findings by Rice and Hoffman (2015) that showed similar age-related quadratic growth in receptive vocabulary in both typically developing children and adolescents. Also worth noting is that, by utilizing a script that includes limited prompting, the lack of examiner scaffolding in our narrative language-sampling procedure may have influenced the total number of utterances produced, such that they were fewer than may be elicited by other methods. This difference may have affected lexical diversity in particular, and thus, we chose to also report our lexical diversity analyses with the inclusion of the total number of utterances variable, which did not alter this pattern of results.

Importantly, beyond informing developmental work, our findings suggest that the language-sampling technique utilized in the Abbeduto et al. narrative procedure is appropriate for measuring change in expressive language abilities in populations whose developmental ages correspond to the ages included in this study (i.e., 4–21 years). This is exciting, given the need for outcome measures that can capture abilities across various language domains without the barriers posed by traditional standardized norm-referenced assessments. Although we did not observe any floor effects in our sample, the youngest age was 4 years, so caution should be observed when utilizing this technique in individuals at younger chronological or developmental ages. Previous studies, however, have successfully used the Abbeduto et al. procedure in individuals with developmental ages as young as 3 years, whose parents reported that they could spontaneously use two- or three-word phrases at least some of the time (Finestack & Abbeduto, 2010; Keller-Bell & Abbeduto, 2007; Kover et al., 2012). Furthermore, the inclusion of older participants in this sample and our findings of cross-sectionally measured age-related change in expressive vocabulary and syntax from early childhood through adolescence extends the potential application of these outcome measures to individuals with intellectual and developmental disabilities at higher developmental levels.

The potential applications of the Abbeduto et al. narrative procedure also extend to populations with other communication disorders without intellectual disability, at least through the chronological age of 18 (i.e., the asymptote of syntactic and lexical performance). At the same time, however, our results emphasize that the particular metrics to be derived from the narrative samples in a clinical trial should be carefully selected with reference to the ages and developmental levels of the participants included in the trial.

Regarding sex differences, we did not observe any significant differences between male and female participants in any of the aspects of expressive language beyond the variance attributed to age. There was, however, a marginally significant interaction term between sex and age for lexical diversity, suggesting that expressive vocabulary may vary slightly between male and female participants at different ages, without any consistent difference across the trajectory. Although our finding is not precise regarding the ages in which there may be some sex differences in expressive vocabulary, we can hypothesize from prior literature suggesting that most sex differences lessen with age and are not consistently evident beyond the preschool years in typically developing children (Bornstein et al., 2004), although Rice and Hoffman (2015) recently demonstrated sex differences through late childhood for receptive vocabulary. Interestingly, though, once controlling for amount of talk (i.e., total number of utterances produced in the narrative) in the model of lexical diversity, the interaction term between sex and age was no longer marginally significant. Regardless, sex differences remain an important consideration when examining performance in populations with intellectual and developmental disabilities, particularly those with differing prevalence rates across sex (e.g., fragile X syndrome, autism spectrum disorder) and suggested differences in underlying neurobiology (Nordahl et al., 2015) or language phenotypes (Abbeduto et al., 2003; Murphy & Abbeduto, 2007; Pierpont et al., 2011).

The Abbeduto et al. narrative procedure is a particularly valuable tool for evaluating individuals with intellectual disabilities, who are vulnerable to floor effects on standardized assessments of spoken language and whose asynchronous profiles of language abilities require analysis across multiple domains (e.g., syntactic complexity, lexical diversity; Abbeduto et al., 2016). The potential of this narrative procedure to capture clinically meaningful gains in abilities over time in a naturalistic context makes it particularly attractive for use as an outcome measure in treatment studies. Its potential as an outcome measure is further enhanced by the data provided in this study, which describe the breadth of age ranges, and thus its corresponding developmental levels, for which it can be used in typically developing individuals. This information is applicable to individuals with communication disorders because it allows for developmentally appropriate matching of the derived expressive language variables to the developmental levels of samples included in clinical trials and other treatment studies. In addition to indexing abilities in the microstructural language domains targeted in this study, this narrative procedure may further be utilized to capture other aspects of language at the macrostructural level of discourse, such as story sequencing (Channell, McDuffie, Bullard, & Abbeduto, 2015) or use of internal state references and other inferential language (Ashby, Channell, & Abbeduto, 2017), although the extent to which it can index such abilities requires further examination and such variables require hand coding rather than computerized computation.

The current study is not without its limitations, including the cross-sectional nature of the age-related developmental trajectories, which limit additional developmental interpretations (i.e., within-subject change over time). The cross-sectional sample is, however, still informative, particularly for interpreting the primary aim of this study—to examine the versatility of the Abbeduto et al. narrative procedure in a typically developing sample. Nonetheless, future efforts utilizing a longitudinal design will strengthen our understanding of how abilities in each aspect of spoken language change with age and should also explore correlates with other measures of expressive language. Although we did not observe clear sex differences in performance in our sample, other demographic variables (e.g., socioeconomic status) may also contribute to individual differences and should be explored in future work, especially given that maternal education was higher in our sample than in the general U.S. population. In addition, because our participants were recruited at one location in the southeastern United States, future studies aiming to provide normative data would need to draw a broader sample and examine the impact of dialect and regional differences on the trajectories of the language variables computed.

Furthermore, it is important for future research to develop more comprehensive descriptive metrics of communicative competence (e.g., story structure, topic maintenance) from narratives derived from this adapted narrative procedure. Because these require hand coding, they extend beyond the scope of gross variables that can be automatically computed for increased usability in clinical trials; nevertheless, they would inform developmental work. In addition, because our sample was more dispersed in the upper age range, more research is needed to replicate our findings in a larger adolescent sample and to determine how performance based on the Abbeduto et al. narrative procedure used in the current study compares with that of other discourse contexts (e.g., more complex narrative or expository discourse). Different discourse contexts place different demands on participants, and storytelling may be more simplistic than other contexts, such as exposition (Nippold et al., 2005; Scott & Windsor, 2000). Although it is reasonable to expect more complex discourse contexts to elicit more advanced language in an older sample, a strength of the context utilized in the current study is its broad applicability to populations with intellectual disabilities that span these chronological ages with highly variable ability levels, including a lower range in which complex discourse may be too challenging.

Finally, the social and cognitive demands of the Abbeduto et al. narrative procedure may pose an additional challenge when applied to certain populations with comorbid social anxiety (e.g., fragile X syndrome) or short-term memory limitations (e.g., Down syndrome). For example, socially, during narration, the burden is placed upon the individual to take the lead role of the speaker, whereas both communicative partners share the role of speaker during a back-and-forth conversational exchange. Cognitively, narration also increases the burden on the speaker to recall past events, organize their representation (e.g., chronological sequences) in the mind, and plan the language needed to communicate them. The current adapted narrative procedure, however, is structured to minimize some of these issues through standardized methods (e.g., providing a warm-up exposure period with the communicative partner, using a picture book to provide memory support), and previous studies have reported evidence for the validity of its use in these populations (Berry-Kravis, Doll, et al., 2013; Kover et al., 2012). In addition, the Abbeduto et al. narrative procedure is intended to capture communicative competence in a naturalistic context that is likely representative of everyday abilities, which also include such social and cognitive demands. Nevertheless, co-occurring symptoms that may interfere with task demands should always be considered when applying this technique to clinical populations. In summary, by providing cross-sectional age-related data on typically developing individuals across a wide age range, the current study provides a critical next step toward increasing the utility of this expressive language sampling technique for individuals with communication disorders, particularly those with intellectual and developmental disabilities.

Acknowledgments

This research was supported by NIH grants R01 HD055345 (PI: Conners) and U54HD079125 (PI: Abbeduto). The authors would like to express their gratitude first to the families who participated in this study and also to the numerous staff who helped throughout this project, from recruitment and scheduling to data collection and transcription.

References

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

Abbeduto, L., Kover, S. T., & McDuffie, A. (2012). Studying the language development of children with intellectual disabilities. In E. Hoff (Ed.), *Handbook of child language research methods* (pp. 330–346). West Sussex, UK: Wiley-Blackwell.

Abbeduto, L., McDuffie, A., Thurman, A., & Kover, S. T. (2016). Language development in individuals with intellectual and developmental disabilities: From phenotypes to treatments. In R. M. Hodapp & D. Fidler (Eds.), *International review of research in developmental disabilities* (pp. 71–118). New York, NY: Academic Press.

Abbeduto, L., Murphy, M., Cawthon, S., Richmond, E., Weissman, M., Karadottir, S., & O'Brien, A. (2003). Receptive language skills of adolescents and young adults with mental retardation: A comparison of DS and fragile X syndrome. *American Journal on Mental Retardation, 108,* 149–160.

A Safety Study of NNZ-2566 in Patients With Fragile X Syndrome. (2013–2017). Retrieved from https://clinicaltrials.gov/ct2/show/NCT01894958 (Identification No. NCT01894958)

Ashby, S. A., Channell, M. M., & Abbeduto, L. (2017). Inferential language use by youth with Down syndrome during narration. *Research in Developmental Disabilities. 71,* 98–108.

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

Berry-Kravis, E., Doll, E., Sterling, A., Kover, S. T., Schroeder, S. M., Mathur, S., & Abbeduto, L. (2013). Development of an expressive language sampling procedure in fragile X syndrome: A pilot study. *Journal of Developmental Behavioral Pediatrics, 34,* 245–251. https://doi.org/10.1097/dbp.0b013e31828742fc

Berry-Kravis, E., Hessl, D., Abbeduto, L., Reiss, A. L., Beckel-Mitchener, A., & Urv, T. K. (2013). Outcome measures for clinical trials in fragile X syndrome. *Journal of Developmental Behavioral Pediatrics, 34,* 508–522. https://doi.org/10.1097/dbp.0b013e31829d1f20

Bishop, D. V. M. (2003). *The Test for Reception of Grammar, Version 2 (TROG-2)*. London: The Psychological Corporation.

Bornstein, M. H., Hahn, C., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language, 24,* 267–304. https://doi.org/0.1177/0142723704045681

Budimirovic, D. B., Berry-Kravis, E., Erickson, C. A., Hall, S. S., Hessl, D., Reiss, A. L., . . . Kaufmann, W. E. (2017). Updated report on tools to measure outcomes of clinical trials in fragile X syndrome. *Journal of Neurodevelopmental Disorders, 9,* 14. https://doi.org/10.1186/s11689-017-9193-x

Channell, M. M., McDuffie, A. S., Bullard, L. M., & Abbeduto, L. (2015). Narrative language competence in children and adolescents with Down syndrome. *Frontiers in Behavioral Neuroscience, 9,* 283. https://doi.org/10.3389/fnbeh.2015.00283

Chapman, R. S., & Hesketh, L. J. (2000). Behavioral phenotype of individuals with Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews, 6,* 84–95.

Craig, H. K., & Washington, J. A. (2000). An assessment battery for identifying language impairments in African American children. *Journal of Speech, Language, and Hearing Research, 43*(2), 366–379.

Dunn, L., & Dunn, L. (2007). *Peabody Picture Vocabulary Test–Fourth Edition.* Bloomington, MN: Pearson Assessments.

Esbensen, A. J., Hooper, S. R., Fidler, D., Hartley, S., Edgin, J., d'Ardhuy, X. L., . . . Urv, T. (2017). Outcome measures for clinical trials in Down syndrome. *American Journal on Intellectual and Developmental Disabilities, 122,* 247–281.

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. https://doi.org/1092-4388/04/4706-1301

Finestack, L. H., & Abbeduto, L. (2010). Expressive language profiles of verbally expressive adolescents and young adults with Down syndrome or fragile X syndrome. *Journal of Speech, Language, and Hearing Research, 53,* 1334–1348.

Ganthous, G., Rossi, N. F., & Giacheti, C. M. (2017). Oral narrative of individuals with fetal alcohol spectrum disorder. *CoDAS, 29,* e20170012. https://doi.org/10.1590/2317-1782/20172017012

Hagerman, R. J., & Hendren, R. L. (Eds.). (2014). *Treatment of neurodevelopmental disorders: Targeting neurobiological mechanisms*. New York, NY: Oxford University Press.

Heilmann, J., Miller, J. F., Inglesias, A., Fabiano-Smith, L., Nockerts, A., & Digney-Andriacchi, K. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders, 28,* 178–187.

Heilmann, J. J., Miller, J. F., & Nockerts, A. (2010). Using language sample databases. *Language, Speech, and Hearing Services in Schools, 41,* 84–95. https://doi.org/10.1044/0161-1461(2009/08-0075)

Heilmann, J. J., Rojas, R., Iglesias, A., & Miller, J. F. (2016). Clinical impact of wordless picture storybooks on bilingual narrative language production: A comparison of the 'Frog' stories. *International Journal of Language & Communication Disorders, 51,* 339–345.

Heilmann, J. J., & Westerveld, M. F. (2013). Bilingual language sample analysis: Considerations and technological advances. *Journal of Clinical Practice in Speech-Language Pathology, 15,* 87–93.

Hessl, D., Nguyen, D. V., Green, C., Chavez, A., Tassone, F., Hagerman, R. J., . . . Hall, S. (2009). A solution to limitations of cognitive testing in children with intellectual disabilities: The case of fragile X syndrome. *Journal of Neurodevelopmental Disorders, 1,* 33–45. https://doi.org/10.1007/s11689-008-9001-8

Hoffman, L. M. (2009). The utility of school-age narrative microstructure indices: INMIS and the proportion of restricted utterances. *Language, Speech, and Hearing Services in Schools, 40,* 365–375. https://doi.org/10.1044/0161-1461(2009/08-0017)

Johnston, J., Miller, J., Tallal, P., & Curtiss, S. (1993). Conversations with children who are language impaired: Asking questions. *Journal of Speech and Hearing Research, 36,* 973–978.

Keller-Bell, Y. D., & Abbeduto, L. (2007). Narrative development in adolescents and young adults with fragile X syndrome. *American Journal on Mental Retardation, 112,* 289–299.

Kover, S. T., McDuffie, A., Abbeduto, L., & Brown, W. T. (2012). Effects of sampling context on spontaneous expressive language in males with fragile X syndrome or Down syndrome. *Journal of Speech, Language, and Hearing Research, 55,* 1022–1038.

Leadholm, B., & Miller, J. (1992). *Language sample analysis: The Wisconsin guide*. Milwaukee, WI: Wisconsin Department of Instruction.

Loban, W. (1976). *Language Development: Kindergarten through Grade Twelve.* Urbana, IL: National Council of Teachers of English.

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities, 32,* 419–436.

Mayer, M. (1973). *Frog on His Own.* New York, NY: Dial Books for Young Readers.

Mayer, M. (1974). *Frog Goes to Dinner.* New York, NY: Dial Books for Young Readers.

McDuffie, A., Thurman, A. J., Channell, M. M., & Abbeduto, L. (2017). Language disorders in children with intellectual disability of genetic origin. In R. Schwartz (Ed.), *Handbook of child language disorders* (2nd ed., pp. 52–81). New York, NY: Taylor & Francis.

Mervis, C. B., & John, A. E. (2010). Cognitive and behavioral characteristic of children with Williams syndrome: Implications for intervention approaches. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 154C,* 229–248.

Miller, J. F. (1991). *Research on child language disorders: A decade of progress; a volume marking the Tenth Anniversary of the Wisconsin Symposium for Research on Child Language Disorders*. Austin, TX: Pro-Ed.

Miller, J., & Iglesias, A. (2006). *Systematic Analysis of Language Transcripts* (SALT Version 9) [Computer Software]. Madison, WI: Language Analysis Lab, University of Wisconsin-Madison.

Miller, J. F., & Klee, T. (1995). Computational approaches to the analysis of language impairment. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 545–572). Oxford, United Kingdom: Blackwell.

Miller, J. F., Long, S., McKinley, N., Thormann, S., Jones, M., & Nockerts, A. (2005). *Language sample analysis II: The Wisconsin guide*. Madison, WI: Wisconsin Department of Public Instruction.

Mills, M. T., Mahurin-Smith, J., & Steele, S. C. (2017). Does rare vocabulary use distinguish giftedness from typical development? A study of school-age African American narrators. *American Journal of Speech-Language Pathology, 26,* 511–523.

Morris-Friehe, M. J., & Sanger, D. D. (1992). Language samples using three story elicitation tasks and maturation effects. *Journal of Communication Disorders, 25*(2–3), 107–124.

Murphy, M. M., & Abbeduto, L. (2007). Gender differences in repetitive language in fragile X syndrome. *Journal of Intellectual Disability Research, 51,* 387–400. https://doi.org/10.1111/j.1365-2788.2006.00888.x

Nippold, M. A. (2010). *Language sampling with adolescents.* San Diego, CA: Plural.

Nippold, M. A., Frantz-Kaspar, M. W., Cramond, P. M., Kirk, C., Hayward-Mayhew, C., & MacKinnon, M. (2014). Conversational and narrative speaking in adolescents: Examining the use of complex syntax. *Journal of Speech, Language, and Hearing Research, 57,* 876–886. https://doi.org/10.1044/1092-4388(2013/13-0097)

Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research, 48,* 1048–1064. https://doi.org/1092-4388/05/4805-1048

Nordahl, C. W., Iosif, A., Young, G. S., Perry, L. M., Dougherty, R., Lee, A., . . . Amaral, D. G. (2015). Sex differences in the corpus callosum in preschool-aged children with autism spectrum disorder. *Molecular Autism, 6,* 26. https://doi.org/10.1186/s13229-015-0005-4

Pierpont, E. I., Richmond, E. K., Abbeduto, L., Kover, S. T., & Brown, W. T. (2011). Contributions of phonological and verbal working memory to language development in adolescents with fragile X syndrome. *Journal of Neurodevelopmental Disorders, 3,* 335–347. https://doi.org/10.1007/s11689-011-9095-2

Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech and Hearing Research, 40*(3), 556–566.

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment (SLI): A longitudinal study from 2½ to 21 years of age. *Journal of Speech, Language, and Hearing Research, 58,* 345–359. https://doi.org/10.1044/2015_JSLHR-L-14-0150

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research, 49*(4), 793–808.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research, 53,* 333–349.

Roid, G., & Miller, L. (1997). *Leiter International Performance Scale–Revised.* Wood Dale, IL: Stoelting.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43,* 324–339. https://doi.org/10.1044/jslhr.4302.324

U.S. Census Bureau. (2015). *Educational attainment. 2011–2015 American community survey 5-year estimates*. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_5YR_S1501&src=pt

Wagner, C. R., Nettelbladt, U., Sahlén, B. B., & Nilholm, C. C. (2000). Conversation versus narration in pre-school children with language impairment. *International Journal of Language and Communication Disorders, 35*(1), 83–93.

Wechsler, D. (1991). *WISC-III: Wechsler Intelligence Scale for Children–Third Edition.* San Antonio: The Psychological Corporation.

Appendix

Prompting Hierarchy for Narrative Language-Sampling Procedure: Part 2 (Story Generation)

Conditions used	Examiner script
FIRST PAGE:	
All participants	*How does the story start?*
ALWAYS WAIT 5–7 SECONDS AFTER CHILD FINISHES TALKING (OR AFTER PREVIOUS PROMPT) BEFORE PROMPTING	
If No Response or *"I don't know,"*	
PROMPT A	*What's happening in this part of the story?*
If Minimal Response,*	
PROMPT B	*That's a good start. Tell me a little more about what's happening in this part of the story.*
If Minimal Response* to Prompt A or Prompt B,	
PROMPT C	*What about the boy? What's he doing?*
If Minimal Response to Prompt C, continue as needed	*What about the frog? What's he doing?* *What about the dog? What's he doing?* *What about the turtle? What's he doing?*
If Minimal Response* to any parts of Prompt C	*Anything else?*
If No Response to above prompts	*Okay. Here's the next page…*
If participant tries to turn the page	*Remember, it's my job to turn the pages, and it's your job to tell me the story.*
SUBSEQUENT PAGES:	
ALWAYS WAIT 5–7 SECONDS AFTER CHILD FINISHES BEFORE PROMPTING/TURNING PAGE	
All participants:	
Each page, after adequate response	*Here's the next page…*
If No Response or *"I don't know,"* **PROMPT A**	*What's happening in this part of the story?*
If No Response or *"I don't know"* to Prompt A, **PROMPT B**	*Tell me everything that is happening in this part of the story + (point).*
If No Response or *"I don't know"* to Prompt B, **PROMPT C1**	*Okay. Here's the next page.*
If Minimal Response* to Prompt B, **PROMPT C2**	*Anything else?* (use sparingly) or *Next page*
LAST PAGE:	
All participants	*How does the story end?*

*Minimal Response is defined as less than adequate. An adequate response (thus not requiring further prompting) meets any of the following: refers to more than one character or action (e.g., "*The boy is getting dressed. And his pets are there.*"), describes an interaction (e.g., "*The animals are looking at the boy*"), or conveys upcoming events in the story (e.g., "*He is getting ready to go out.*").