Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s10803-018-3628-4

COMMENTARY

Commentary: Measuring Language Change Through Natural Language Samples

Mihaela Barokova[1] · **Helen Tager-Flusberg**[2]

Abstract
The role of language in autism spectrum disorder (ASD), more specifically, its function in social communication and strong predictive power on future outcomes, warrants language assessments that have good psychometric properties that capture the heterogeneity of language ability found among diagnosed individuals. Given the rapid growth in intervention and treatment research, there is an urgent need for the development and implementation of outcome measures that are easily obtained and sensitive to change. In this commentary, we argue for the use of natural language samples as measures of expressive language and communication for this purpose and review the literature on their implementation in ASD research. Conceptual and measurement issues are discussed and future developments are outlined.

Keywords Autism · Language · Outcome measures · Natural language samples

Introduction

Although language is no longer a defining criterion for the diagnosis of autism spectrum disorder (ASD; American Psychiatric Association 2013), it remains of great importance for research purposes and clinical practice because of its central role in development. From the earliest stages, language serves two primary goals—communication with others and transmission of information. Later in development language plays an integral role in the development of higher order cognitive processes including theory of mind and executive functions (e.g., de Villiers 2007; Hale and Tager-Flusberg 2003). In the context of ASD, the emergence of expressive language before the age of five or six has been identified as one of the strongest predictors of positive long-term outcomes for diagnosed children (e.g., Howlin et al. 2004; Venter et al. 1992) and is inversely related to atypical behaviors (Dominick et al. 2007). Because of its multifaceted role in development and predictive power on future outcomes, language has become an important target of ASD interventions (e.g., Kasari et al. 2010; Tager-Flusberg and Kasari 2013). Yet, measuring change in language ability still presents a significant challenge in ASD research and has been limited by relying on standardized tests or parent report measures. In this commentary we argue that measures derived from natural language samples (NLSs) hold the potential of useful outcome measures because they can capture improvement in language ability across the wide range of age and language levels that may characterize participants in clinical trials and can provide a more comprehensive view of language than other types of measures.

We begin with a review of language and the different ways of assessing language. We then turn to studies using NLSs in ASD to characterize language in this, and related disorders, and then focus on their use to obtain outcome measures in a small number of behavioral intervention studies. We conclude with a discussion of the further research needed to establish the psychometric properties of NLS-derived outcome measures and with guidelines on how they could be incorporated in future clinical trials.

✉ Mihaela Barokova
barokova@bu.edu

Helen Tager-Flusberg
htagerf@bu.edu

[1] Center for Autism Research Excellence, Department of Psychological and Brain Sciences, Boston University, 100 Cummington Mall, Boston, MA 02215, USA

[2] Boston University, 100 Cummington Mall, Boston, MA 02215, USA

Published online: 05 June 2018

Language in Typical Development and Autism Spectrum Disorder

Language is a system of communication governed by a set of rules. It consists of structural and pragmatic components, which follow separate, though intertwined developmental pathways. Phonology is the domain of language concerned with speech sounds and how and in what linguistic contexts they can be combined. The basic unit of phonology, the phoneme, is defined as a contrastive sound in a language. Phonemes combined together make up morphemes, which are the smallest meaningful units of speech. Morphology is the domain of language concerned with how morphemes can be combined to form and change the meaning of words. Words, in turn can be grouped into phrases and sentences. The set of rules that govern this highly systematic process comprise the syntax, or grammar, of the language. The study of the meaning of the formed sentences and their key building blocks, words, constitutes the semantics of the language.

Phonology, morphology, syntax, and semantics are the structural components of language that allow us to transmit information. However, they do not function independently because the language they comprise is used in specific contexts to serve specific communicative needs. Pragmatics is the component concerned with how context contributes to meaning, and, more generally, how language is used as a tool for communication. It encompasses language and other non-linguistic signals (e.g., prosody, gesture, facial expression) and includes rules for conversations and other forms of discourse, as well as the functional roles of utterances (e.g., request, reprimand, comment, etc.) as used in a range of contexts. Thus, the ability to use and integrate context with the structural components of language places pragmatics at the heart of social communication. Here it is important to identify the distinction between spoken language, expressive language, and expressive communication. Spoken language refers to the speech produced by individuals. Most research in the field of typical development and in autism has focused on the study of spoken language. Expressive language includes spoken language, as well as speech produced with the help of augmentative and alternative communication (AAC) devices. Studying expressive language is particularly important in autism considering the increasing popularity of AAC devices and their inclusion in autism interventions (e.g., Kasari et al. 2014). Finally, expressive communication is an umbrella term, which includes both verbal communication, i.e., expressive language, and nonverbal communication, for example, facial expressions, gestures, etc. Further efforts to better understand expressive communication in autism are also warranted (e.g., Grossman et al. 2013; Medeiros and Winsler 2014).

The successful acquisition of the structural and pragmatic components of language involves a set of hierarchically organized skills as reflected in expressive communication. Consider, for example, collecting a language sample from a child playing with their mother. The child's production of speech sounds reflects their speech planning ability and phonological development. The ability to combine morphemes into words and words into phrases reveals the level of mastery of morphology and syntax, respectively. The integration of gestures and facial expressions with verbal language to serve a communicative goal, and the child's conversational turn taking represent that child's pragmatic skills and reflect their degree of social engagement. These examples illustrate how a sample of expressive language and communication can be a versatile source of information about that child. Much of what we know about typical language development comes from the collection and analyses of expressive language samples (e.g., Brown 1973). Furthermore, developmental milestones within and across language domains have been firmly established for typically developing children (e.g., Gleason 2017; Oller et al. 2014), thus providing the means for comparing the spoken language skills of children with different language disorders, including ASD, to those of typically developing children.

The fact that expressive communication provides information about a speaker's mastery of the major domains of language and communication makes its assessment especially suitable for individuals at all ages with ASD (Kjelgaard and Tager-Flusberg 2001; Lord et al. 2004). Although the structural components of language appear to be unaffected in some individuals with ASD, the majority acquire spoken language but are impaired relative to their peers, and up to one-third fail to acquire spoken language skills beyond a minimal level (Kim et al. 2014). Research suggests that, in general, there are parallels between language development in ASD and in typical or other language disordered children (Kim et al. 2014). In contrast, the pragmatics domain is deeply embedded in social context, thus individuals with ASD whose core impairments are in the social domain, exhibit a wide range of impairments including problems using nonverbal communicative cues, deictic expressions (Hobson et al. 2010), responding contingently in conversations (Tager-Flusberg and Anderson 1991), or reconstructing the sequence of events in a narrative (Diehl et al. 2006) all of which impact their ability to engage with other people. For this reason, pragmatic impairments are included in one of the criteria for social communication deficits under the current DSM 5 definition of ASD (American Psychiatric Association 2013).

In ASD, there is not only a wide range of language abilities found among speakers across the spectrum, but also in individual speakers' mastery across language domains. Thus, individuals with ASD often present with a mixed

language profile, suggesting they are at different stages in their development across the different language domains (Tager-Flusberg 1994; Tager-Flusberg et al. 2009), and in receptive and expressive language (Weismer et al. 2010). We focus in this commentary primarily on expressive language as it is more reliably assessed and is typically the target of intervention studies, though we note that studies generally do not include non-speech aspects of expressive language (Tager-Flusberg et al. 2009; Tager-Flusberg 2000).

Measuring Expressive Language Change in ASD

Perhaps the most significant potential use of expressive language measures is in evaluating changes that are the result of development over time or as a consequence of treatment. Recognizing the potential of expressive language as an outcome measure for treatment studies, Tager-Flusberg and her colleagues, who formed a Working Group under the direction of the National Institutes of Health, proposed a developmental framework for the measurement and reporting of expressive language of individuals with ASD across different domains of language (Tager-Flusberg et al. 2009). The framework identifies spoken language benchmarks to be used in determining the language ability of children and a common terminology for its reporting to allow for the comparison of outcomes across treatment and intervention studies. In addition, Tager-Flusberg et al. (2009) advise the use of a combination of different types of measures (NLSs, standardized assessments, and parent report) for the comprehensive evaluation of expressive language in individuals with ASD; a recommendation that has also been made for other neurodevelopmental disorders such as Fragile X syndrome (FXS) (Berry-Kravis et al. 2013b).

ASD interventions and clinical trials often enroll heterogeneous samples, including participants across a wide range of age and/or language ability (e.g., Siller et al. 2013; Bent et al. 2012). Interventions can be broadly classified as either long-term or short-term based on their duration and may target core ASD symptoms, language, or other co-occurring symptoms. With regards to their outcomes, significant developmental changes are to be expected as a result of longer-term interventions. For example, Estes et al. (2015) reported gains in language, social functioning, and IQ in young children after receiving a 2-year-long in-home intensive comprehensive early intervention (the Early Start Denver Model) that were maintained and had generalized to other aspects of functioning at a 2 year follow up after the intervention had ended. In contrast, interventions shorter in duration would not be expected to yield the same extensive gains. Nevertheless, many different kinds of interventions might lead to changes in participants' spoken language that reflect improvements in social-communication, language, repetitive behaviors, or even challenging behaviors. For example, since expressive language is inversely related to the severity of atypical behaviors in ASD (Dominick et al. 2007), interventions that reduce these behaviors may also result in improvements in language use. Thus, measures of language that capture both significant developmental changes as well as those that are smaller but functionally meaningful, have potential for use as outcome measures. What is most important is that language outcome measures should be psychometrically sound and meet established reliability and validity standards.

When measuring change in spoken language, clinical trials typically use standardized tests or parent questionnaires. Although these assessments are appropriate for investigating language phenotypes in ASD, they are often less useful as comprehensive and sensitive outcome measures. Standardized language tests (e.g., Peabody Picture Vocabulary Test; Dunn and Dunn 2007; or the Comprehensive Assessment of Spoken Language; Carrow-Woolfolk 2017) provide direct assessments of language and have the advantage of possessing empirically validated psychometric properties. However, care must be taken to ensure that the norming sample for standardized tests included individuals from different clinical populations, including ASD. With the exception of vocabulary tests, standardized language tests typically have a limited age range; thus, participants varying widely in age and language ability could not be tested on the same outcome measure. There are also limited options for measuring changes in pragmatic skills using standardized tests since many aspects of this domain cannot, by definition, be captured in a structured testing context. Because test scores have a standard error of estimation, shorter duration clinical trials may lie within the standard error of estimation of, for example, age-equivalent scores, thereby limiting the ability of such scores to capture genuine changes in spoken language. Floor effects might further conceal changes for individuals with ASD whose spoken language ability falls outside the range of ability for which the test was designed. Other factors including motivation or anxiety can further influence performance; consequently, test scores might misrepresent an individual's spoken language abilities. This is especially relevant for individuals with ASD whose social and communicative impairments may directly influence their test performance (Tager-Flusberg 2000). Finally, many language tests should only be administered by trained experts and can take a significant amount of time to administer, which could make them a less feasible option for some clinical trials.

Information about a child's spoken language and its change over time can also be obtained from parent questionnaires such as MacArthur-Bates Communication Development Inventories (Fenson et al. 2007) or the Children's Communication Checklist 2nd Edition (Bishop 2003), both

of which have been validated for use with children with ASD (e.g., Nordahl-Hansen et al. 2014). Parent measures provide valuable information about children's language use in everyday situations as reported by someone who spends considerable time with them, in contrast to the administration of a standardized test by a stranger, often in an unfamiliar environment. Nevertheless, one important limitation of using parent questionnaires as outcome measures in clinical trials is that they may be subject to placebo effects. For example, in one striking randomized controlled trial, Guastella et al. (2015) found that parents who believed their children received oxytocin reported greater improvements post-treatment compared to parents who believed their children received the placebo, when in fact, standardized assessments indicated no actual improvement for participants in either condition. In this study, parents reported on their children's social responsiveness, developmental behaviors, and emotional problems, nevertheless, it is likely that parents reporting on changes in language could be subject to similar placebo effects.

An alternative to standardized tests and parent questionnaires is NLS analysis. NLSs are recordings of spontaneous expressive language that can be elicited in different contexts (e.g., free play, conversation, narration) and collected from different people (e.g., parent, examiner) in different settings (e.g., clinic, school, home, lab). Once collected, these language samples can be analyzed and coded for structural and pragmatic features, which makes them especially useful in ASD research considering the within and cross language domain heterogeneity that is characteristic of the disorder.

NLSs are a rich source of information about language phenotypes in autism, and they are excellent candidates for outcome measures in clinical trials. For example, they can be collected from participants spanning a wide range of age and language ability, which allows for the comparison of outcomes of heterogeneous samples of participants enrolled in clinical trials. The same procedures for eliciting language samples can be used at multiple time points over the course of an intervention trial without running into the risk of practice effects (Abbeduto 2017). Furthermore, NLSs capture the everyday speech of the speakers in contrast to standardized tests, and thus reflect potential improvements of functional importance to everyday social interactions, which are often targets of ASD intervention studies. Measures derived from NLSs may also be more sensitive to change in speech compared to standardized assessments and parent report measures, which is another advantage of using them as outcome measures (e.g., Casenhiser et al. 2013; Kaiser and Roberts 2013). Since they are collected in more naturalistic settings, NLSs are often more fun for the participant compared to standardized tests. Finally, they have the advantage of requiring little training for their administration and can be administered and coded by naïve examiners, blinded to which arm of a trial the child is assigned and when (pre- or post-intervention) the NLS was collected.

Natural Language Samples in ASD Research

Although NLSs are good candidates for outcome measures in intervention studies, research using measures derived from NLSs has primarily focused on describing the language phenotypes of the disorder (e.g., Chiang 2009; Colle et al. 2008; Kover and Abbeduto 2010). Nevertheless, given the parallels in the methods for collecting and analyzing language samples for both types of investigations, the research describing the language phenotype in ASD can be used as a stepping-stone in the process of their selective utilization as outcome measures. In this commentary, we review a sample of studies from the field of autism and other neurodevelopmental disorders that use NLSs in their methodology and either assess and describe language ability or report on a randomized control trial. The specific articles were selected to represent a wide range of NLS collection methods and NLS-derived measures.

We present here, and in Table 1, an overview of how NLSs have been used in the assessment of the expressive language phenotype of individuals with ASD and related neurodevelopmental disorders, including FXS. We focus on the defining features of NLSs including the characteristics of the target speakers, the context, setting, conversational partner, length of samples, the measures derived from them, and their potential use in the context of clinical trials. Table 2 provides a quantitative summary of how these NLS characteristics are distributed across the reviewed studies.

Target Speaker Characteristics

NLSs have been used to evaluate the expressive language of individuals with neurodevelopmental disorders covering a broad range of ages and language ability. For example, NLSs have been collected from toddlers with ASD as young as 9 months of age (Patten et al. 2014) through adolescents (Chiang 2009) and adults (e.g., Colle et al. 2008; McCabe et al. 2013). Within a single study, NLSs were compared across participants with ASD between the ages of 2 and 16 (Chiang 2009). Berry-Kravis et al. (2013a) collected language samples from individuals with FXS spanning an even wider age range 5–35 years. Furthermore, NLSs have been used to assess the expressive language of individuals with ASD with different levels of productive language: from children with fewer than five functional words and use AAC devices (e.g., Chiang 2009) to children who use phrase speech (Klusek et al. 2014) to adults who can narrate a story (Colle et al. 2008). These examples illustrate how NLSs can be applied to heterogeneous participant samples,

Table 1 Summary of the purpose, participant characteristics, standard and spontaneous expressive language measures and findings of studies relying on natural language sampling procedures

References	Study; design	N	Subjects	Age (years)	Standard Language Measures	NLS—setting, context, partner, length	NLS measures	Main findings
Berry-Kravis et al. (2013a)	Develop a sampling procedures for expressive language; test–retest 2–4 weeks apart	36	36 × FXS	M = 18 5–35	None	Narrative and conversation Researcher Approx. 17 min in total	MLU NDW TNU Proportion of complete and intelligible C-units with mazes Proportion of fully or partially unintelligible C-units	All NLS measures derived from both elicitations contexts were reliable Minimal practice effects were reported MLU and NDW were correlated with VABS scores Proportion of unintelligible C-units and C-units with mazes were less strongly correlated with VABS scores TNU was not correlated with VABS scores
Chiang (2009)	Examine and compare the spontaneous and elicited communication in individuals with ASD with limited spoken language (some used aided and unaided AAC); at one time	34	17 × Taiwanese ASD 17 × Australian ASD	M = 7.7 2.2–16.2	None	Naturalistic setting and context at home and at school Whoever was around 120 min	Coded for Spontaneous vs. elicited communication Use and type of AAC Function of CA Type of communicative partner Types of partner's responses Type of activity	The amount of spontaneous expressive communication was higher than elicited expressive communication Spontaneous expressive communication was positively correlated with age and negatively correlated with autism severity Differences in preference of type of communication and function of CA were observed between elicited and spontaneous communication

Table 1 (continued)

References	Study; design	N	Subjects	Age (years)	Standard Language Measures	NLS—setting, context, partner, length	NLS measures	Main findings
Colle et al. (2008)	Examine and compare general narrative abilities and specific pragmatic abilities between adults with high-functioning autism or Asperger Syndrome and typically developing controls	24	12 × ASD 12 × TD	M = 27.5 M = 27.2	None	Narrative	TNU Coded for Episodes of the story Kinds of referential expressions Temporal relations and mental state expressions	Participants with ASD used fewer referential and temporal expressions, and personal pronouns compared to TDs
Condouris et al. (2003)	Examine the relationship between spontaneous speech measures and standardized measures in ASD; at one time	44	44 × ASD	4–14	PPVT-III EVT CELF-P CELF-3	Play Parent 30 min	MLU NDWR IPSyn	MLU correlated with CELF grammar subscores NDWR correlated with CELF lexical-semantic subscores and PPVT-III and EVT scores IPSyn only correlated with other spontaneous measures and underestimated the language ability of ASD participants
Hale and Tager-Flusberg (2005a)	Examine whether rate of noncontingent utterances is related to ASD symptomatology; at one time	57	57 × ASD	4–13.92	PPVT-III EVT	Play Parent 30 min	Coded for Noncontingent and contingent responses Imitation	The proportion of noncontingent utterances was significantly correlated with ASD symptom severity
Hogan-Brown et al. (2013)	Compare narrative abilities of children with different neurodevelopmental disorders; at one time	94	20 × ASD 23 × FXS + ASD 18 × FXS 17 × DS 16 × TD	ASD M = 9.04 FXS + ASD M = 10.28 FXS M = 9.73 DS M = 11.01 TD M = 4.55	EVT/PPVT EVT PPVT	Narrative Researcher	Microstructure MLU Number of propositions/utterances Frequency and number of different types of utterances Macrostructure Number main story episodes Frequency of inappropriate, irrelevant and unintelligible utterances	No group differences in narrative length, MLU, and syntactic complexity Interaction between diagnosis, nonverbal cognitive ability, and complex syntax use

Table 1 (continued)

References	Study; design	N	Subjects	Age (years)	Standard Language Measures	NLS—setting, context, partner, length	NLS measures	Main findings
Kover and Abbeduto (2010)	Compare the expressive language of adolescents with developmental disorders across narrative and conversational contexts; several sessions over 1 or 2 days	44	8 × FXS + ASD 20 × FXS 16 × DS	FXS + ASD M = 16.38 FXS M = 15.61 DS M = 16.05	None	Narrative and conversation Researcher Approx. 17 min in total	MLU NDWR Number of C-units attempted per min Proportion of utterances with mazes Proportion of intelligible utterances	Overall, participants used more different words during the conversation and their speech exhibited higher syntactic complexity during the narrative context FXS group was more talkative than DS group During the narrative context FXS and FXS + ASD groups were more talkative than DS Overall, participants' speech was more dysfluent during the conversation context

Table 1 (continued)

References	Study; design	N	Subjects	Age (years)	Standard Language Measures	NLS—setting, context, partner, length	NLS measures	Main findings
Kover et al. (2014)	Compare the expressive language produced by children with ASD during ADOS and during play with a parent and during play with an examiner Examine whether children get classified differently based on the benchmarks framework; at one time during second visit	63	63 × ASD	M = 3.75	MSEL PLS-4	ADOS (first 15 min) and play with researcher and with parent 30 min in total	TNU Percentage of intelligible utterances NDW MLU Number of requests, comments, and turn-taking	All of the variables differed across sampling context; Highest TNU for the parent–child interaction followed by the examiner–child interaction followed by the ADOS context Percentage of intelligible utterances and NDW were higher for the parent– and examiner–child interactions than for the ADOS MLU was higher for the examiner–child interaction than for the parent–child interaction and the ADOS More requests were made during the parent–child interaction than during the examiner–child interaction and the ADOS More comments were made during the parent– and examiner–child interactions than during the ADOS More turns were taken during the parent–child interaction than during the ADOS Overall, participants were categorized into lower developmental language phases for the domains of phonology, vocabulary, and pragmatics when using the ADOS as a sampling context compared to the other contexts

Table 1 (continued)

References	Study; design	N	Subjects	Age (years)	Standard Language Measures	NLS—setting, context, partner, length	NLS measures	Main findings
Kover et al. (2012)	Examine the effects of sampling context on expressive language; at one time	57	27 × FXS 15 × DS 15 × TD	FXS and DS 10–17 TD 3–6	None	Narrative and conversation Researcher 11–21 min	TNU Number of complete and intelligible utterances Number of attempted utterances per minute MLU Total number of lexical word roots Proportion of utterances containing mazes Proportion of utterances partially or fully unintelligible	TNU was higher for the conversation than the narrative context Number of attempted utterances per minute for the DS group was higher for the conversation context than the narrative context MLU was highest for the TD group followed be the FXS group followed by the DS group TD group exhibited higher lexical diversity than the FXS and DS groups The groups differed in how fluent they were across elicitation contexts FXS and TD groups were more intelligible than the DS group The autism symptom severity of participants with FXS was related to TNU Researcher's behavior was found to be related to the participants' performance
McCabe et al. (2013)	Compare personal narratives between adults with ASD and typically developing adults	34	16 × ASD 18 × TD	ASD M = 20.7 18–26 TD M = 21.3 18–27	NART	Narrative through conversational procedure Researcher 10–20 min	High Point (HP) analysis of narrative structure Topic maintenance Event sequencing Informativeness Narrative Assessment Protocol Referencing Conjunctive cohesion Dysfluencies	The personal narratives of participants with ASD were characterized by less complex structure than TDs' but did not differ in length (in words or propositions)

Table 1 (continued)

References	Study; design	N	Subjects	Age (years)	Standard Language Measures	NLS—setting, context, partner, length	NLS measures	Main findings
Murphy and Abbeduto (2007)	Compare verbal perseveration of male and female adolescents with FXS across language sample contexts; at one time during second visit	24	24 × FXS 8 × female 16 × male	Female M = 15.35 Male M = 17.03	OWL	Narrative and conversation Researcher > 10 min	Coded for Utterance-level repetition Topic repetition Conversational device repetition Proportion of utterance-level repetition Proportion of topic repetition Proportion of conversational device repetition Proportion of within-utterance repetition	More topic repetition was observed for the conversation context than the narrative context for both males and females Gender differences were observed only for the conversational context with males producing higher proportion of conversational device repetition than females
Patten et al. (2014)	Compare the rates of canonical babbling and volubility of infants who were later diagnosed with ASD to those who were not; Retrospective video (two 5-min video segments) analysis at two time points	37	23 × later diagnosed with ASD 14 × TD	Time 1: 0.9–1.0 Time 2: 1.3–1.6	MSEL VABS	Home Naturalistic context and conversational partners Approx. 10 min	Ratio of canonical to total number of syllables Counted canonical and noncanonical syllables Syllables per minute Coded for Speech-like vocalizations	TD participants were more likely to be categorized to have reached the canonical babbling stage than ASD participants at both time points TD had higher canonical babbling ratio than ASD participants at both time points; ratio was positively correlated with age TD group had higher volubility than ASD at both time points; effect of age too Volubility and canonical babbling status were a predictor of later diagnosis

Table 1 (continued)

References	Study; design	N	Subjects	Age (years)	Standard Language Measures	NLS—setting, context, partner, length	NLS measures	Main findings
Schoen et al. (2011)	Examine the vocalizations of toddlers with ASD and age- and language-matched controls	64	30 × ASD 11 × TDA 23 × TDL	ASD: 1.6–3.0 TDA: 1.6–3.0 TDL: 0.11–1.1	ASD MSEL VABS ADOS-G CSBS-DP Controls VABS MSEL VABS TDA: CSBS-DP TDL: ADOS-T	For TDL (5 min parent–child interaction) Or for ASD and TDA CSBS (15 min clinician–child interaction) Approx. 20 min	Singleton consonant inventory SSL Coded for Speechlike vs. non-speech For ASD and TDA Consonant blend inventory Word production inventory PCC	Participants with ASD produced more atypical non-speech vocalizations than both TDA and TDL ASD produced more squeals compared to TDA and TDL TDA group produced more consonants than ASD 48% of ASD and all of TDA participants produced consonant blends TDA group's vocalizations exhibited more complex syllable structure than ASD and TDL

Table 1 (continued)

References	Study; design	N	Subjects	Age (years)	Standard Language Measures	NLS—setting, context, partner, length	NLS measures	Main findings
Suh et al. (2014)	Examine the pragmatic domain of the language of HFA, OO, and TD children by evaluating their narratives; during testing: 2–3 visits, but the sample is from a single session	45	15 × HFA 15 × OO 15 × TD	9.3–15.8	WASI	Narrative (ADOS Mod 3 Tuesday book) Examiner	TNW TNU MLU NDW/TNU Number story elements Proportion of ambiguous pronouns Number of repetitions, self-corrections, fillers Mental and emotional state expressions, causal references Presence of idiosyncratic language, unusual references, and character naming	No difference in narrative length and lexical diversity between groups Number of story elements mentioned was lower for HFA group than TD group Proportion of ambiguous pronoun references was higher for HFA than TD and OO Number of repetitions and self-corrections were lower for TD than HFA and OO HFA and OO more likely than TD to have idiosyncratic language in their narratives TD were more likely than HFA to name the characters while narrating

Only significant results are reported

AAC augmentative and alternative communication, *ADOS* Autism Diagnostic Observation Schedule, *C-units* communicative units, *CA* communication acts, *CELF* Clinical Evaluation of Language Fundamentals (P-preschool), *CSBS* Communication and Symbolic Behavior Scales, *DS* Down syndrome, *EVT* Expressive Vocabulary Test, *FXS* Fragile X syndrome, *HFA* high functioning autism, *IPSyn* index of productive syntax, *MLU* mean length of utterance, *MSEL* Mullen Scales of Early Learning, *NART* National Adult Reading Test, *NDW* number of different words, *NDWR* number of different word roots, *OO* optimal outcome, *OWL* Oral and Written Language scales, *PCC* percent consonants correct, *PLS* Preschool Language Scale, *PPVT* Peabody Picture Vocabulary Tests, *TD* typically developing, *TDA* typically developing age-matched, *TDL* typically developing language-matched, *TNU* total number of utterances, *TNW* total number of words, *VABS* Vineland Adaptive Behavior Scales, *WASI* Wechsler Abbreviated Scale of Intelligence

Table 2 Quantitative summary of the characteristics of the NLSs used in the studies included in Table 1

NLS characteristic	Characteristic subcategories	Number of studies out of 14
Speaker's age	Infant and toddler (< 2 years)	2
	Preschooler (2–4 years)	4
	Middle childhood (4–12 years)	6
	Adolescence (12–18 years)	6
	Adulthood (> 18 years)	3
Context	Play	4
	Conversation	4
	Narrative	8
	Semi-structured/other assessment	3
Setting	Lab	12
	Home	2
	School/kindergarten	1
Conversational partner	Researcher	10
	Parent	4
	Whoever was around	2
Length	10–20 min	7
	20–30 min	2
	> 30 min	1

Note that many studies have collected NLSs from a wide range of participants or relied on several contexts and/or conversational partners. Also many studies do not report on the duration or setting of their language samples

which in turn allows for a comparison of improvement in language across participants while using the same NLS-derived measures.

Context, Setting, and Conversational Partner

The most popular context in which NLSs have been collected is play with developmentally appropriate toys, especially for younger participants, and these samples are usually elicited in conversation with a parent or examiner (e.g., Condouris et al. 2003; Hale and Tager-Flusberg 2005a; Kaiser and Roberts 2013; Kover et al. 2014). This elicitation approach can be implemented in the home, clinic, school or research lab. Other studies use naturalistic recordings, which include routine activities the participants engage in on a daily basis (e.g., Chiang 2009) or use home videos (Patten et al. 2014) thus including as conversational partners different family members, peers, or even teachers, which means that the language samples are likely to be representative of the language used daily by the participant.

Other studies use narratives as the context for eliciting NLSs (e.g., Hogan-Brown et al. 2013; Kover and Abbeduto 2010; Kover et al. 2012; Suh et al. 2014). This context is more suitable for older children, whose language skills (including vocabulary and grammatical knowledge) are sufficiently advanced to satisfy the demands required for the narration of a story. Studies describing expressive language in FXS have reported that language used in narratives is characterized by more complex syntactic structures compared to conversational interactions. However, conversations elicit more lexically diverse utterances (Kover and Abbeduto 2010) and more utterances overall (Kover et al. 2012). These qualitative and quantitative differences in expressive language contingent on the elicitation procedure warrant caution in the choice of sampling context.

Other contexts that have been used for the collection of NLSs are the semi-naturalistic free play and conversational segments of standardized assessments including the Autism Diagnostic Observation Schedule (ADOS) or the Communication and Symbolic Behavior Scales (e.g., Klusek et al. 2014; Kover et al. 2014; Paul et al. 2013; Schoen et al. 2011). In the context of these more structured assessments (Klusek et al. 2014; Kover et al. 2014) or during intervention sessions (Kasari et al. 2014), NLSs, although convenient to collect, might not be fully representative of the participant's expressive language. For example, Kover et al. (2014) compared the characteristics of the language samples elicited from young children with ASD varying both the communicative partner (parent or examiner), and context (play session or ADOS). Overall frequency of speech differed significantly across these contexts with the highest number of utterances and different words produced during parent–child free play, followed by examiner–child free play, and the fewest elicited during the ADOS. Pragmatic differences were also found: children made more requests and comments, as well as took more turns during free play than during the ADOS. Kover et al. (2014) also categorized their participants into language phases according to Tager-Flusberg et al.'s (2009) benchmark framework separately for each language sample collected (parent–child, examiner–child, and ADOS). In the domains of phonology, vocabulary, and pragmatics, participants were categorized into lower language phases during the ADOS compared to free play thus illustrating the clinical implications associated with using different elicitation contexts to evaluate the language ability of individuals with ASD.

Overall, the results of this study suggest that NLSs collected by parents during naturalistic interactions may be optimal, at least for younger children with ASD. Nevertheless, the choice of sampling context should be guided by the goals of the research, the language and developmental level of the participants, and the properties of the protocol. For studies of older children, Abbeduto and colleagues developed a standardized approach for the collection of language samples from individuals with FXS that has excellent psychometric properties (e.g., Berry-Kravis et al. 2013a) and

they have begun examining its use with individuals with other neurodevelopmental disorders including ASD and Down syndrome with promising results (Abbeduto 2017; Kover and Abbeduto 2010).

Length

There are two general guidelines typically followed when determining the length of language samples: the minimum number of utterances or minimum amount of time required to yield reliable and valid measures. When assessing the language ability of typically developing children, some studies have provided recommendations based on the age/developmental stage of the child, the specific measure in question, or the sampling context (e.g., Heilmann et al. 2010). There is, however, wide variation in the duration of language samples used in the assessment of expressive language of TD children and individuals with ASD. The length of the language sample is usually determined by the aims of the study and the sampling context. Home recordings tend to be longer, for example, up to 2 hours (Chiang 2009), compared to lab-based free play samples, which average around 30 min (e.g., Hale and Tager-Flusberg 2005a). Studies assessing expressive language in FXS typically collect conversation samples that are 10 min long as well as narratives that vary in duration based on the language ability of the speaker (e.g., Abbeduto et al. 1995).

Even though different sample lengths in terms of minimum number of utterances or duration have been used across different studies, the optimal length or duration, one that is long enough to yield reliable and valid measures, has yet to be empirically determined for participants with ASD across different ages and elicitation contexts.

Measures Derived from Natural Language Samples

Measures derived from NLSs in the context of characterizing the language phenotype of ASD (summarized in Table 1) or used as outcome measures in intervention studies (summarized in Table 3) cover the different domains of expressive language from phonology to pragmatics.

Phonological measures can capture the number of different phonemes produced to assess early milestones of speech development (e.g., Schoen et al. 2011). Utterance planning and fluency have also been evaluated using NLS-derived measures such as the proportion of intelligible utterances (e.g., Kover et al. 2014).

Several different measures are used to evaluate morphosyntactic aspects of expressive language. Mean length of utterance in morphemes (MLUm) or words (MLUw) is a measure of grammatical complexity used to define stages of language development for typically developing children (Brown 1973) and is significantly related to the grammatical complexity of the expressive language of children with ASD with MLU of up to three morphemes (Scarborough et al. 1991). It has been widely used as a measure in ASD (e.g., Casenhiser et al. 2015; Condouris et al. 2003; Hale and Tager-Flusberg 2005b; Hogan-Brown et al. 2013; Kaiser and Roberts 2013; Kasari et al. 2014; Klusek et al. 2014; Kover et al. 2014; Suh et al. 2014) and FXS research (e.g., Kover and Abbeduto 2010; Kover et al. 2013). MLU correlated with standardized test subscores relating to grammatical development on the Clinical Evaluation of Language Fundamentals (CELF; Semmel et al. 1995) in a group of 4- to 14-year old children with ASD (Condouris et al. 2003) providing evidence for the construct validity of this NLS-derived measure. The Index of Productive Syntax (IPSyn) (Scarborough 1990), which captures the occurrence of a range of morphosyntactic constructions and is correlated with MLU for children with ASD (Scarborough et al. 1991), has also been used to measure grammatical complexity for children with ASD (Condouris et al. 2003; Kaiser and Roberts 2013). However, Condouris et al. (2003) reported that for verbally fluent children with ASD, unlike MLU, IPSyn was not correlated with standardized test scores and underestimated grammatical ability when compared to other measures calling into question the validity of this measure.

The number of different word roots (NDWR) is a measure of lexical diversity that is often used in studies of ASD (e.g., Condouris et al. 2003; Kaiser and Roberts 2013; Kasari et al. 2014; Suh et al. 2014) or FXS (e.g., Kover and Abbeduto 2010). Although different variants of this measure have been used including total NDWR (Kasari et al. 2014), number of different words (Kaiser and Roberts 2013), number of different words over total number of words (Suh et al. 2014), and word production inventory (Schoen et al. 2011), they all capture the richness of a speaker's expressive vocabulary and are likely to be highly correlated with one another. NDWR was strongly correlated with standardized vocabulary test scores (Peabody Picture Vocabulary Test and Expressive Vocabulary Test) and the lexical-semantic subscores on the CELF, in a sample of children with ASD (Condouris et al. 2003), providing evidence for the good construct validity of the measure.

NLS-derived measures of pragmatic skills are, perhaps, the most important use of NLSs. One language sample can capture many different aspects of this domain. For example, NLSs have been used to classify different speech acts by coding the function of utterances such as comments (e.g., Kasari et al. 2014), or acknowledgements (e.g., Kover et al. 2014). How well a speaker tailors their speech to the characteristics and knowledge of the communicative partner can be examined by coding for the use and clarity of deictic expressions (e.g., personal pronouns; Colle et al. 2008). The ability to tell a narrative has been evaluated by coding the NLS for narrative plot elements included, character

Table 3 Summary of the purpose, participant characteristics, NLS characteristics and measures, and findings of intervention and treatment studies relying on NLS-derived outcome measures

References	Study; design	N	Subjects	Age (years)	NLS—setting, context, partner, length	NLS measures	Main findings
Casenhiser et al. (2015)	Comparison of MEHRIT to community treatment; RCT	51	51 × ASD	2.0–4.11	Play Parent 25 min	MLU TNU Filled pauses Coded types of CA based on their function	MEHRIT group made greater gains measured in TNU, MLU, number of different CAs, and rate of responding to partner compared to the community treatment group
Deitchman et al. (2010)	Examine the effects of incorporating video feedback into self management training on social initiation; multiple-probe design	3 (2)	3 × ASD 2 × TD	1 × 5 1 × 6 1 × 7	School Daily school activities 30 min	Frequency of verbal Social initiations (e.g., introducing a New idea, commenting, requesting, expressing emotions, trying to attract attention) to peers	Significant increase in the frequency of verbal social initiations post-intervention
Kaiser and Roberts (2013)	Comparison of therapist-only to parent- and therapist-implemented EMT; randomized group design	77	16 × ASD 18 × DS 44 × DD	M = 3.39 2.5–4.5	Play Adult 30 min	NDW IPSyn MLU Number of unique targets % of child utterances containing targets	At 6 months: parent + therapist EMT group had longer MLU and higher NDW compared to therapist-only EMT during trained activity MLU, NDW, use of targets from story at 6 and 12 months for parent condition were all higher compared to the therapist-only condition No significant differences between groups in standardized scores after intervention but still increase compared to pre-intervention
Kasari et al. (2014)	Comparison of JASP + EMT to JASP + EMT + SGD; SMART	61	61 × minimally verbal with ASD	5–8	Play Researcher 20 min	TSCU TDWR TCOM PSCU WPM MLU NUWC	SGD use led to greater TSCU, TDWR, and TCOM at 24 weeks All measures were higher for SGD group at 12 weeks
Mohammadzaheri et al. (2014)	Comparison of the effects of structured ABA and PRT on language and communication; RCT	30	30 × ASD	6–11	School Describe 6 pictures Researcher	MLU	PRT group improved significantly more in MLU from pre- to post-intervention compared to ABA group

Table 3 (continued)

References	Study; design	N	Subjects	Age (years)	NLS—setting, context, partner, length	NLS measures	Main findings
Paul et al. (2013)	Comparison of MCT + parent responsiveness training to RMIA + parent responsiveness; quasi-experimental design	22	22 × minimally verbal with ASD	RMIA: 4.3 MCT: 3.5	CSBS Play 20 min	Number of words produced	Both RMIA and MCT groups produced more words post-treatment and had achieved higher CDI scores as reported by their parents; These results were maintained for 3–6 month No change in VABS-Expressive language scale scores pre- to post-treatment First expressive language stage achieved according to Tager-Flusberg et al.'s (2009) framework: 50% of RMIA group and 42% of MCT group Children with higher joint attention scores on CSBS benefited more regardless of treatment

Only significant results are reported

ABA applied behavior analysis, *CA* communication acts, *CDI* Child Development Inventory, *CSBS* Communication and Symbolic Behavior Scales, *DD* developmental delay, *DS* Down syndrome, *EMT* enhanced milieu teaching, *IPSyn* index of productive syntax, *JASPER* Joint Attention Symbolic Play Engagement and Regulation, *MCT* milieu communication training, *MEHRIT* Milton & Ethel Harris Research Initiative treatment, *MLU* mean length of utterance, *NDW* number of different words, *NUWC* number of unique word combinations, *PRT* pivotal response treatment, *PSCU* proportion of socially communicative utterances, *RCT* randomized control trial, *RMIA* rapid motor imitation antecedent, *SGD* speech-generating device, *SMART* sequential multiple assignment randomized trial, *TCOM* total number of comments, *TD* typically developing, *TDWR* total number of different word roots, *TNU* total number of utterances, *TSCU* total number of spontaneous communicative utterances, *WPM* words per minute

naming, and unusual references among others (e.g., Suh et al. 2014). Conversational skills have been studied by coding for the rate of contingent and noncontingent responses (e.g., Hale and Tager-Flusberg 2005a) or topic-related utterances (Tager-Flusberg and Anderson 1991). Furthermore, the content of the speech like repetition of topics has been evaluated by determining the proportion of topic repetitions in the language sample (e.g., Murphy and Abbeduto 2007). Other pragmatic deficits associated with ASD such as speech perseveration, echolalia, use of neologisms, or inappropriate use of deictic expressions can also easily be coded and analyzed from a NLS.

Numerous ASD treatment studies have used NLS-derived pragmatic outcome measures. For example, in their study evaluating the effectiveness of a social initiation training, Deitchman et al. (2010) found differences in the frequency of verbal social initiations including commenting, requesting, trying to attract attention, and introducing a new idea to peers pre- and post-training. Casenhiser and colleagues found intervention-related increases in total number of utterances with different communicative purpose and in rate of responding to one's communicative partner (Casenhiser et al. 2015). Still another intervention study found differences in the total number of comments when comparing the outcomes of different training programs (Kasari et al. 2014). In these examples, NLS-derived measures that have been used to describe pragmatic ability in autism were used as outcome measures by investigating change from pre- to post-intervention. They have primarily focused on assessing the ability of children with autism to use utterances with different communicative functions in different contexts. However, change in other aspects of pragmatics could be evaluated just as easily from NLSs.

One advantage of using NLS-derived outcome measures is their sensitivity to change. Thus, a number of studies have reported no change as a result of intervention in standardized language scores but a significant change in NLS-derived language measures of children with ASD (e.g., Casenhiser et al. 2013; Kaiser and Roberts 2013). These results demonstrate how the choice of language outcome measure could affect the evaluation of the effectiveness of a specific intervention and favor the use of NLS-derived measures. For example, Kaiser and Roberts (2013) compared the effects of two interventions (parent + therapist Enhanced Milieu Training (EMT) vs. therapist only EMT) on the communication of children with intellectual disabilities of different etiology (ASD, Down syndrome, and developmental delay). The two intervention groups did not differ post-intervention on any of the standardized assessments. However, one group, parent + therapist EMT intervention, made more gains in MLU and NDWR post-intervention as measured in a NLS.

In another study reporting on the effectiveness of a social-interaction-based ASD intervention, Casenhiser et al. (2013) found that the intervention group and the control group made similar gains on standardized test scores. However, later analyses of NLSs from the participants found that on measures including MLU, number of different communicative acts, rate of responding to partner, and contingency of responses the intervention group made more gains than the control group (Casenhiser et al. 2015).

These studies along with those discussed earlier that validate NLS-derived measures against standardized tests have begun to pave the way for the use of NLS-derived outcome measures in ASD intervention research.

NLS-Derived Language Outcome Measures: Guidelines and Future Directions

Before using NLS-derived outcome measures in ASD intervention and clinical trial studies, a number of conceptual and measurement issues still need to be addressed. These can be divided into two main categories: ones that concern the establishment of standard elicitation protocols and ones that concern the evaluation of the psychometric properties of the measures derived from such protocols.

Standard Elicitation Protocol

Extensive research has demonstrated that the context, setting, and conversational partner used in the collection of language samples all lead to systematic differences in the amount and kind of language being elicited. Therefore, it is essential that equivalent elicitation procedures are followed for the collection of the pre- and post-samples in clinical trials (and, when possible, at regular time intervals during the trial to provide richer information on change over time). The key here is to follow equivalent though not necessarily identical procedures in order to avoid the risk of practice effects. The contexts (free play, conversation, narrative) should be consistent, but the specific toys, conversation topics, or narrative elicitation method (e.g., book) can be different, if they have been shown to elicit similar language samples. Furthermore, the utilization of a standardized elicitation protocol across different intervention studies or clinical trials will allow for a more rigorous comparison of the effectiveness of different interventions.

Two key characteristics for this protocol to meet are: good test–retest reliability, i.e., no practice effects of the protocol and good inter-tester administration fidelity, i.e., the protocol is administered consistently in the same manner across testers and time as to not introduce changes that might affect the quality and quantity of the elicited language.

Efforts in the field have already been made to create a standardized elicitation protocol across neurodevelopmental disorders and wide ranges of ability (e.g., Abbeduto 2017).

Abbeduto and colleagues (2017) have designed a protocol that includes a semi-structured conversation and story narrative to be used with children and adolescents with FXS, ASD, and Down syndrome. Their NLS has good test–retest reliability, and its NLS-derived measures have been cross-validated with standardized measures of language and communication. This protocol is a great starting point and a promising direction for researchers in the field of autism. Yet, further research efforts are necessary to expand this protocol to include a more naturalistic free play context and a familiar conversational partner and to adapt it to be suitable for much younger children and older children and adults with more limited verbal abilities.

Outcome Measures

Outcome measures derived from NLSs to be used in ASD clinical trials should also possess good psychometric properties including construct validity, ability to detect change, reliability, and validity. Furthermore, the extent to which language-based outcome measures reflect core ASD symptoms including social communicative impairments or restricted interests and repetitive behaviors should be explored further. While some progress has been made validating structural and pragmatic measures derived from language samples (e.g., Abbeduto 2017; Condouris et al. 2003; Hale and Tager-Flusberg 2005a, b), these studies have focused on children with fluent (though often impaired) language. Future research should aim to assess the construct validity of NLS-derived measures for individuals that reflect the entire spectrum of language ability in ASD, including minimally and low verbal children, adolescents and adults by examining their relationship with already-established standardized assessment measures or parent report measures that reflect the construct of interest.

A key measurement issue to be taken into account in the choice of NLS-derived outcome measures is their ability to detect change, as has been demonstrated in a few studies (e.g., Casenhiser et al. 2013; Kaiser and Roberts 2013). Nevertheless, an important question remains as to what constitutes a *meaningful change*, one that has real-life implications for the individual. One way to address this is to establish the associations between changes in outcome scores and changes in clinical judgment and/or parent- or self-reported daily functioning, which has yet to be done in ASD research.

The reliability and validity of NLS-derived measures also needs to be examined further with a focus on ensuring that the psychometric properties of the measures are consistent for males and females, and across the range of age and language ability of participants that could be enrolled in clinical trials. This can be accomplished with studies carefully designed to assess the psychometric properties of the NLS-derived measures by following procedures long employed to psychometrically validate standardized assessments and parent reports. Furthermore, with the increased implementation of intervention and clinical trials in countries beyond North America, more research is needed to validate NLS-derived measures for languages other than English and to systematically examine the role of the specific cultural context on the qualitative and quantitative characteristics of the NLS-derived measures.

Once future research establishes the validity of NLS-derived outcome measures, the specific choice of which ones to use in clinical trials should ultimately be driven by their applicability to use with the target population, their ability to detect meaningful change, and their potential to address the research questions of interest in the most informative and cost-effective way.

Conclusion

Language outcome measures derived from NLSs are excellent candidates for evaluating change in many aspects of language and communication and for comparing the effectiveness of interventions across studies and participant groups. NLSs can be collected and outcome measures evaluated from individuals with ASD across the entire spectrum of the disorder following the same sampling procedure. The choice of specific outcome measures can be easily tailored to the goals of the specific project. A single language sample depending on the level of analysis can measure change in language ability, including pragmatic skills, as well as in other core symptoms of ASD.

Although a promising and potentially better alternative to standardized assessments, NLS-derived outcome measures have yet to be widely adopted in ASD research. Before this happens, research needs to determine the best practices for their implementation. Language sampling guidelines need to be established (e.g., context, communicative partner, duration) with potentially developing and adopting a standard elicitation protocol across studies. Recommendations need to be made about the coding and computation of the sample-derived measures. Their validation against standardized assessments and parent reports is also necessary. Even though further research on the psychometric properties of NLS-derived measures is warranted, based on what has already been done in the field, we argue that NLSs have the potential to play an instrumental role in ASD treatment and intervention research.

Author Contributions Both authors conceived of the manuscript. Mihaela Barokova carried out the systematic review guided by HTF and drafted the manuscript. Both authors worked on revising successive drafts.

Funding This paper was supported by a research grant from the Simons Foundation Autism Research Initiative, #383660, to Dr. Helen Tager-Flusberg.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abbeduto, L. (2017). *Expressive language sampling as an outcome measure: Psychometric properties across neurodevelopmental disorders*. Presentation presented at the Simons Foundation Autism Research Initiative Novel Outcome Measures Workshop, New York City, NY.

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*(5), 279–288.

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

Bent, S., Bertoglio, K., Ashwood, P., Nemeth, E., & Hendren, R. L. (2012). Brief report: Hyperbaric oxygen therapy (HBOT) in children with autism spectrum disorder: A clinical trial. *Journal of Autism and Developmental Disorders, 42*(6), 1127–1132.

Berry-Kravis, E., Doll, E., Sterling, A., Kover, S., Schroeder, S., Mathur, S., & Abbeduto, L. (2013a). Development of an expressive language sampling procedure in fragile X syndrome: A pilot study. *Journal of Developmental and Behavioral Pediatrics, 34*(4), 245–251.

Berry-Kravis, E., Hessl, D., Abbeduto, L., Reiss, A., Beckel-Mitchener, A., & Urv, T. (2013b). Outcome measures for clinical trials in fragile X syndrome. *Journal of Developmental and Behavioral Pediatrics, 34*(7), 508–522.

Bishop, D. V. M. (2003). *The children's communication checklist—2*. London: Psychological Corporation.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Carrow-Woolfolk, E. (2017). *Comprehensive assessment of spoken language, second edition (CASL-2) [Manual]*. Torrance, CA: Western Psychological Services.

Casenhiser, D. M., Binns, A., McGill, F., Morderer, O., & Shanker, S. G. (2015). Measuring and supporting language function for children with autism: Evidence from a randomized control trial of a social-interaction-based therapy. *Journal of Autism and Developmental Disorders, 45*(3), 846–857.

Casenhiser, D. M., Shanker, S. G., & Stieben, J. (2013). Learning through interaction in children with autism: Preliminary data from a social-communication-based intervention. *Autism, 17*(2), 220–241.

Chiang, H.-M. (2009). Differences between spontaneous and elicited expressive communication in children with autism. *Research in Autism Spectrum Disorders, 3*(1), 214–222.

Colle, L., Baron-Cohen, S., Wheelwright, S., & Van Der Lely, H. K. J. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders, 38*(1), 28–40.

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech and Language Pathology, 12*(3), 349–358.

De Villiers, J. (2007). The interface of language and theory of mind. *Lingua. International Review of General Linguistics, 117*(11), 1858–1878.

Deitchman, C., Reeve, S. A., Reeve, K. F., & Progar, P. R. (2010). Incorporating video feedback into self-management training to promote generalization of social initiations by children with autism. *Education and Treatment of Children, 33*(3), 475–488.

Diehl, J. J., Bennetto, L., & Young, E. C. (2006). Story recall and narrative coherence in high-functioning autism spectrum disorders. *Journal of Abnormal Child Psychology, 34*(1), 87–102.

Dominick, K. C., Ornstein, N., Lainhart, J., Tager-flusberg, H., & Folstein, S. (2007). Atypical behaviors in children with autism and children with a history of language impairment. *Research in Developmental Disabilities, 28*, 145–162.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test–Fourth edition*. San Antonio, TX: The Psychological Corporation.

Estes, A., Munson, J., Rogers, S., Greenson, J., Winter, J., & Dawson, G. (2015). Long-term outcomes of early intervention in 6-year-old children with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 54*(7), 580–587.

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., & Bates, E. (2007). *MacArthur-Bates communicative development inventories-III*. Baltimore: Brooks Publishing.

Gleason, J. (2017). *The development of language* (9th ed.). Boston: Pearson.

Grossman, R. B., Edelson, L. R., & Tager-Flusberg, H. (2013). Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism. *Journal of Speech Language and Hearing Research, 56*(3), 1035.

Guastella, A., Gray, K., Rinehart, N., Alvares, G., Tonge, B., Hickie, I., & Einfeld, S. (2015). The effects of a course of intranasal oxytocin on social behaviors in youth diagnosed with autism spectrum disorders: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 56*(4), 444–452.

Hale, C., & Tager-Flusberg, H. (2005a). Brief report: The relationship between discourse deficits and autism symptomatology. *Journal of Autism and Developmental Disorders, 35*(4), 519–524.

Hale, C., & Tager-Flusberg, H. (2005b). Social communication in children with autism. *Autism, 9*(2), 157–178.

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*(3), 346–359.

Heilmann, J., Nockerts, A., & Miller, J. F. (2010). Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools, 41*(4), 393–404.

Hobson, R. P., García-Peérez, R. M., & Lee, A. (2010). Person-centred (deictic) expressions and autism. *Journal of Autism and Developmental Disorders, 40*(4), 403–415.

Hogan-Brown, A. L., Losh, M., Martin, G. E., & Mueffelmann, D. J. (2013). An investigation of narrative ability in boys with autism and fragile X syndrome. *American Journal on Intellectual and Developmental Disabilities, 118*(2), 77–94.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry, 45*(2), 212–229.

Kaiser, A., & Roberts, M. (2013). Parent-implemented enhanced milieu teaching with preschool children with intellectual disabilities. *Journal of Speech, Language, and Hearing Research, 56*(1), 295–309.

Kasari, C., Gulsrud, A., Wong, C., Kwon, S., & Locke, J. (2010). Randomized controlled caregiver mediated joint engagement

intervention for toddlers with autism. *Journal of Autism and Developmental Disorders, 40*(9), 1045–1056.

Kasari, C., Kaiser, A., Goods, K., Nietfield, J., Mathy, P., Landa, R., … Almirall, D. (2014). Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 53*(6), 635–646.

Kim, S. H., Paul, R., Tager-Flusberg, H., & Lord, C. (2014). Language and communication in autism. In F. Volkmar, S. Rogers, R. Paul & K. Pelphrey (Eds.), *Handbook of autism and pervasive developmental disorders* (Vol. 2, 4th ed., pp. 230–262). Hoboken, NJ: Wiley.

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes, 16*(2–3), 287–308.

Klusek, J., Martin, G., & Losh, M. (2014). A comparison of pragmatic language in boys with autism and fragile X syndrome. *Journal of Speech, Language, and Hearing Research, 57*(5), 1692–1707.

Kover, S. T., & Abbeduto, L. (2010). Expressive language in male adolescents with fragile X syndrome with and without comorbid autism. *Journal of Intellectual Disability Research, 54*(3), 246–265.

Kover, S. T., Davidson, M., Sindberg, H., & Weismer, S. (2014). Use of the ADOS for assessing spontaneous expressive language in young children with ASD: A comparison of sampling contexts. *Journal of Speech, Language, and Hearing Research, 57*(6), 2221–2233.

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*(4), 1022–1038.

Lord, C., Risi, S., & Pickles, A. (2004). Trajectory of language development in autistic spectrum disorders. In M. Rice & S. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 1–38). Mahwah, NJ: Erlbaum. 2004.

McCabe, A., Hillier, A., & Shapiro, C. (2013). Brief report: Structure of personal narratives of adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(3), 733–738.

Medeiros, K., & Winsler, A. (2014). Parent-child gesture use during problem solving in autistic spectrum disorder. *Journal of Autism and Developmental Disorders, 44*(8), 1946–1958.

Mohammadzaheri, F., Koegel, L. K., Rezaee, M., & Rafiee, S. M. (2014). A randomized Clinical trial comparison between pivotal response treatment (PRT) and structured applied behavior analysis (ABA) intervention for children with autism. *Journal of Autism and Developmental Disorders, 44*(11), 2769–2777.

Murphy, M. M., & Abbeduto, L. (2007). Gender differences in repetitive language in fragile X syndrome. *Journal of Intellectual Disability Research, 51*(5), 387–400.

Nordahl-Hansen, A., Kaale, A., & Ulvund, S. E. (2014). Language assessment in children with autism spectrum disorder: Concurrent validity between report-based assessments and direct tests. *Research in Autism Spectrum Disorders, 8*(9), 1100–1106.

Oller, J., Oller, S. D., Badon, & Linda, C. (2014). *Milestones: Normal speech and language development across the life span* (2nd ed.). San Diego: Plural Publishing, Inc.

Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal Patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency. *Journal of Autism and Developmental Disorders, 44*(10), 2413–2428.

Paul, R., Campbell, D., Gilbert, K., & Tsiouri, I. (2013). Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(2), 418–431.

Scarborough, H., Rescorla, L., Tager-Flusberg, H., Fowler, A., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics, 12*(1), 23–45.

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics, 11*, 1–22.

Schoen, E., Paul, R., & Chawarska, K. (2011). Phonology and vocal behavior in toddlers with autism spectrum disorders. *Autism Research, 4*(3), 177–188.

Semmel, E., Wiig, B., & Secord, W. (1995). *Clinical evaluation of language fundamentals-3*. San Antonio: Psychological Corporation, Harcourt Brace.

Siller, M., Hutman, T., & Sigman, M. (2013). A parent-mediated intervention to increase responsive parental behaviors and child communication in children with ASD: A randomized clinical trial. *Journal of Autism and Developmental Disorders, 43*(3), 540–555.

Suh, J., Eigsti, I.-M., Naigles, L., Barton, M., Kelley, E., & Fein, D. (2014). Narrative performance of optimal outcome children and adolescents with a history of an autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders, 44*(7), 1681–1694.

Tager-Flusberg, H. (1994). *Constraints on language acquisition: Studies of atypical children*. Hillsdale: L. Erlbaum Associates.

Tager-Flusberg, H. (2000). The challenge of studying language development in autism. In L. Menn & N. Bernstein Ratner (Eds.), *Methods for studying language production* (pp. 313–332). Mahwah, NJ: Lawrence Erlbaum Associates.

Tager-Flusberg, H., & Anderson, M. (1991). The development of contingent discourse ability in autistic children. *Journal of Child Psychology and Psychiatry, 32*, 1123–1134.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research, 6*(6), 468–478.

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., … Yoder, P. (2009). 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, 52*(3), 643–652.

Venter, A., Lord, C., & Schopler, E. (1992). A follow-up study of high-functioning autistic children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 33*(3), 489–507.

Weismer, S. E., Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism spectrum compared to toddlers with developmental delay. *Journal of Autism and Developmental Disorders, 40*(10), 1259–1273.