

The State of Natural Language Sampling in Autism Research: A Scoping Review

Samantha N Plate 

Abstract

Background and aims: Caregiver reports and standardized assessments have been the primary methods used to study language development in autism. However, these forms of measurement are often coarse, complicated by floor effects and reporter bias, and limited by the fact that they only capture how children *can* use language at a single moment in time, rather than how children *actually* use language during everyday interactions. These limitations have led to recent calls for the use of natural language sampling (NLS) as a fine-grained, developmentally appropriate, and contextually relevant measure of everyday communication. The number of studies using NLS to study language in autism has increased substantially in the last 15 years, resulting in a wide array of sampling methods and measures. Given both the increasing prevalence of NLS methods in the autism literature and the variability in sampling approaches and measures, this scoping review addresses the following questions:

1. What populations have been studied using NLS?
2. Which data collection methods are most prevalent in NLS research?
3. How are measures of language derived from NLS?

Method: Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, a search for studies published in the last 15 years across three databases was conducted. After removing duplicates, 4,671 titles and abstracts were screened and 59 papers met inclusion criteria. Sample characteristics, natural language collection methods, and derived measures were extracted and tabled for each study. The most prevalent NLS methods and measures in autism language research are reviewed and the benefits and drawbacks of various methods are discussed.

Main contribution: This scoping review highlights subgroups of the autistic population that have been underrepresented in NLS studies—in particular, minimally/nonspeaking school-aged autistic children. This article also examines means to collect a “naturalistic” sample of language. Notably, studies did not address whether autistic children exhibit different social communication skills when talking to different types of social partners. Broadly, research has underreported key methodological details, making comparisons across studies difficult.

Conclusions: This review highlights the appropriate use of NLS across development in autism and makes recommendations for NLS future research.

Implications: Additional work is needed to address the gaps described in this article and replicate previous findings to identify patterns of natural language across the literature.

Keywords

Autism, natural language sampling, naturalistic observation

Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

Corresponding Author:

Samantha N Plate, University of Pittsburgh, 210 South Bouquet Street, 3309 Sennott Square, Pittsburgh, PA 15213, USA.
Email: snp52@pitt.edu



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Introduction

Autism spectrum disorder (ASD) is characterized by differences in social communication, and while expressive language challenges are not included in the current diagnostic criteria (American Psychiatric Association, 2013), many autistic children demonstrate language impairments (Tager-Flusberg et al., 2005; Tager-Flusberg & Kasari, 2013). Moreover, delays in language development are among the most common first concerns reported by parents, even before their children receive a formal diagnosis (Herlihy et al., 2015; Kozłowski et al., 2011). Differences and delays in language are predictive of a later autism diagnosis (Ozonoff et al., 2010; Plate et al., 2021; Warlaumont et al., 2014; Yankowitz et al., 2019), and delayed expressive language in particular, is associated with a variety of later outcomes in autistic children. Language samples collected in early childhood (3–6 years) have been shown to predict language, verbal IQ, adaptive behavior, and autism symptoms in adulthood (LeGrand et al., 2021). Similarly, parent-reported vocabulary at ages 2 and 3 predict autism symptoms, as well as cognitive, adaptive, and language abilities, at age 9, in autistic children (Luyster et al., 2007). For these reasons, the development of language and communication has been identified as a key developmental process in the unfolding of autism in children. Thus, a large body of research has focused on describing expressive language skills and measuring developmental changes in autism.

Given this pivotal role of language in the development and characterization of autism, it is critical to identify valid and reliable measures to quantify expressive language in autistic children. Caregiver reports and standardized assessments have been the primary methods for assessing expressive language in children. However, these measures are often coarse and complicated by floor and ceiling effects and reporter bias (Barokova & Tager-Flusberg, 2018; Laghi et al., 2022). Additionally, although some studies have found that caregiver reports of language are consistent with direct assessments (Miller et al., 2017), others have found only moderate correlations with observational language samples, with evidence that caregivers underestimate child language, especially for verbs (Feldman et al., 2005; Kover et al., 2018; Tardif et al., 1999). Moreover, while caregiver reports from two vocabulary checklists completed in the same visit were highly consistent, they were not perfect—caregivers endorsed that their child says a certain word on one form, but not the other for 10% of words (Arunachalam et al., 2022). Different questionnaires may result in under, or over, reporting by caregivers, and it is therefore important to use multiple measures to get accurate estimates of language abilities. Additionally, Fenson et al. (1994) identified that test–retest reliability on a common vocabulary checklist (Fenson et al., 2007) is lower for children who perform below the median. This evidence

suggests that caregiver report measures of expressive language may be less accurate for children with language delays, such as autistic children. Standardized assessments are also limited by the fact that they only capture how a child *can* use language at a single moment in time, but not how the child *actually* uses language during everyday interactions (Kasari et al., 2013; Naples et al., 2022; Tamis-LeMonda et al., 2017). This limitation may be particularly salient for children on the autism spectrum, as their performance on standardized measures may be more variable than in other populations, and they may be more sensitive to novel testing environments. Taken together, this body of work underscores the importance of gathering data from multiple sources (e.g., parent report, direct assessment, and naturalistic observation) to obtain a complete picture of a child’s language abilities, particularly for autistic children.

Natural Language Sampling

These methodological limitations have led to recent calls for the use of natural language sampling (NLS) as a fine-grained, developmentally appropriate, and contextually relevant measure of everyday communication (Barokova et al., 2020). NLS has been defined as “recordings of spontaneous expressive language that can be elicited in different contexts (e.g., free play, conversation, narration) in different settings (e.g., clinic, school, home, lab)” (Barokova & Tager-Flusberg, 2018, p. 4). A limited body of research suggests that NLS measures are reliable and valid (Hudry et al., 2023; McDaniel et al., 2018). One advantage of NLS is its ability to capture various levels of language, including phonetics, syntax, and pragmatics (Barokova & Tager-Flusberg, 2018).

Another strength of NLS methods is that they are more contextually salient and ecologically valid compared to standardized measures of language. Broadly speaking, NLS methods capture language in the natural contexts in which it is typically used, although setting, context, and how “natural” these methods are may vary. Furthermore, the measures derived from these methods are more fine-grained than standard measures such as vocabulary inventories. NLS measures can be used to assess minute changes in behaviors, such as rates of vocalizations, lengths of utterances, and consonant inventories (i.e., the diversity of consonant sounds a child produces; Wetherby et al., 2007; Woynaroski et al., 2016). Therefore, NLS methods provide advantages for studies requiring measures sensitive to change in expressive language (e.g., intervention studies).

NLS methods have been primarily used with infants to measure vocal development before the production of spoken words. A large body of work has described the development of vocalizations and language in prospective studies of infants with elevated likelihood of developing autism (see Jones et al., 2014; Yankowitz et al., 2019 for

reviews). Broadly, these studies analyzed vocalization rates, social directedness (i.e., how much infants direct their vocalizations toward a social partner), canonical babbling (i.e., the combination of consonant and vowel sounds), and acoustics of cries taken from NLS (Yankowitz et al., 2019). These infant studies found that prelinguistic vocalizations, including atypical acoustics, can serve as markers of early developmental differences and predict later diagnosis. However, NLS research beyond prospective infant studies is more varied and has not been well reviewed.

One group for whom NLS may be a more appropriate measure than standardized assessments is individuals with minimal or no spoken language (also referred to as minimally/nonspeaking or minimally/nonverbal). These individuals often score at floor on standardized measures (Barokova & Tager-Flusberg, 2018; Kasari et al., 2013; Trembath et al., 2019), which creates challenges in measuring variability. Additionally, previous work indicates that standard scores, percentiles, and age equivalent scores are not appropriate for measuring meaningful change over time in children with significant delays and disabilities (Perry et al., 2009). This concern has led to recommendations for using raw scores or growth scale values when employing standardized assessments to measure development, and more preferably, using observational methods when possible (Brady et al., 2020; Chen et al., 2024; Farmer et al., 2020; Kasari et al., 2013; Perry et al., 2009). Finally, children with minimal or no spoken language and other challenges (e.g., limited mobility, disruptive behavior) have historically been deemed “untestable.” These challenges suggest that standardized assessments may not accurately reflect the development of skills in the individuals who are most significantly affected by autism and has led to their underrepresentation in language research. To address these issues, Kasari et al. (2013) evaluated measures of expressive language and identified NLS as one of the most appropriate methods for individuals with minimal or no spoken language. For example, children who produce few to no words often produce pre-speech sounds, such as babbles, that may not be measured on a standardized assessment or caregiver report. The flexibility of NLS methods, however, would allow for researchers to collect measures of consonant inventories, canonical babbling ratios, and frequency of vocalizations.

While NLS has been used extensively to study infants, it is less regularly employed as a method to study language in older children and adolescents, especially those who are minimally/nonspeaking. Although previous reviews have identified common sampling approaches and measures in the infant literature, little is known about how research has employed NLS in autistic children and adolescents. Infant language is sampled primarily in the context of caregiver–infant interactions, and the contexts that are developmentally appropriate for capturing

language in older children are more varied. While the field of child language development in autism appears to be moving toward NLS methods, the recent increase in studies using NLS has consequently resulted in a wide array of sampling methods and measures, leading to the current review.

The Current Study

In the past 15 years, the number of studies using NLS to measure the language of autistic children has increased substantially (see Figure 1). Based on both the increasing prevalence of NLS methods in the autism literature and the variability in sampling approaches and measures, a scoping review is needed to describe the state of the field and catalog the participant groups, methods, and measures that have been included in this research to date. This review will focus on studies of diagnosed autistic children older than 18 months of age, excluding studies of NLS in infancy, which have been the focus of prior reviews and meta-analyses (see Jones et al., 2014; McDaniel et al., 2018; Yankowitz et al., 2019). It will address the following set of fundamental questions regarding how NLS methods and measures are used to study language in autism: (a) who is included in NLS research?; (b) how are natural language samples collected?; and (c) how are measures of language derived from NLS?

Who is Included in NLS Research?

A number of studies have used NLS methods to document the emergence of expressive language in preschoolers and the progression of language used for social communication in autistic children and adolescents. This review will examine the prevalence of NLS studies in these age groups. Additionally, it is common for studies to compare autistic children to “typically developing” peers. However, for studies interested in identifying differences that are specific to autism, rather than potentially driven by language or developmental differences more broadly, it may be appropriate to include samples of children with developmental or language delays who do not have autism. This review aims to identify the composition of these participant samples.

It is also critical to evaluate the language level of participants that are recruited for NLS research. Despite the fact that individuals with minimal to no spoken language make up an estimated 15–30% of the autistic population (Hughes et al., 2023; Tager-Flusberg et al., 2005; Tager-Flusberg & Kasari, 2013), and NLS is a recommended measure of expressive language for minimally/nonspeaking children (Kasari et al., 2013), little is known about how it has been used in this population. The current review will identify which subgroups of the autistic population have been included in NLS studies, and which subgroups should be prioritized for future research.

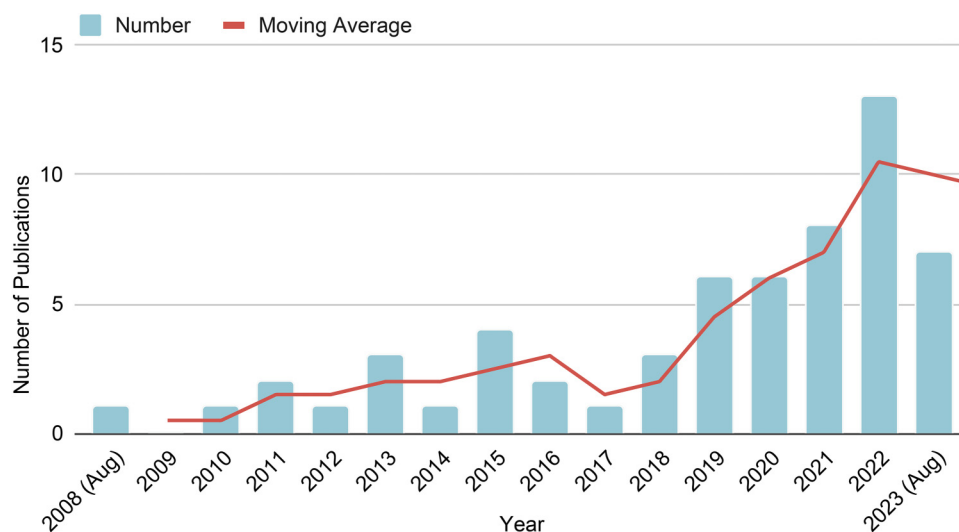


Figure 1. NLS Publications Over Time.

This figure displays the number of publications included in the current review by year published. The red line shows the moving average, taking the average of the current year and prior year.

How are Natural Language Samples Collected?

There is currently no consensus on what constitutes a “natural” language sample. The following contexts have all been described as “naturalistic” in previous work: language produced in the context of a clinical assessment (Cola, Yankowitz et al., 2022; Plate et al., 2021), language produced at home during unstructured time (Parladé & Iverson, 2015; Warlaumont et al., 2014; Winder et al., 2013), conversations with a novel social partner in a research setting (Cola, Zampella et al., 2022; Song et al., 2021), and vocalizations produced during semistructured play with a caregiver in the lab (Hudry et al., 2023). Moreover, relatively few studies address whether autistic children exhibit different social communication skills when talking to children versus adults, autistic versus non-autistic social partners, and familiar versus novel social partners (but see Hilvert et al., 2020; Kover et al., 2014). Additionally, as children develop, their social worlds widen to include peers, teachers, romantic partners, and community members.

When considering using NLS methods, it is important to identify who the “natural” and developmentally appropriate conversational partners are for the population being studied. The current paper will review studies from toddlerhood to late adolescence to identify the most common approaches to collecting NLS samples and review the advantages and limitations of these methods. A developmental framework is used to discuss which contexts are developmentally appropriate for collecting NLS among autistic children. Additionally, this review will examine the degree to which the extant literature captures the natural interactions children have with proximal adults and children in the child’s

immediate environment (e.g., teachers, siblings, peers, caregivers other than mothers).

How are Measures of Language Derived from NLS?

Natural language samples yield a variety of dependent variables that capture different features of communication and language. A recording of child language can be transcribed to classify the types of words children use (Boorse et al., 2019; Cola, Yankowitz et al., 2022), coded to determine rate of vocalization production (Paul et al., 2011; Plate et al., 2021), or analyzed acoustically to quantify variability in pitch (Fusaroli et al., 2017). One major issue in NLS work is whether automated analysis can be used on recorded data or whether human coders are needed to conduct time-intensive transcription and vocalization coding (Yankowitz et al., 2019). Research has suggested that automated vocalization identification, through programs such as Language ENvironment Analysis (LENA; Xu et al., 2009) is less accurate than human coding (Lehet et al., 2021). However, automated coding is more efficient, may still provide a representative measure of an individual’s language, and is more fine-grained than standardized measures.

Recent studies have suggested that human-coded samples are reliable and valid measures of language (Hudry et al., 2023), but validity is often defined by existing standardized measures. Therefore, measurement validity is likely to vary across samples of participants and collection methods. For example, children identified as having minimal spoken language may produce a very small number of words (as measured by a standardized parent questionnaire), but may vocalize frequently in an NLS context,

Table 1. Search Terms for each Key Word.

Key word	Search terms
Natural language (in title or abstract)	"natural language sampling," "natural language," "NLS," ("natural" AND "language"), "Language Environment Analysis," "LENA," "speech," "speak,*" "vocal,*" "conversati,*" "communicat,*" "talkativeness," "interact*"
Autism (in title and abstract)	"autism," "autistic," "spectrum disorder," "Asperger," "pervasive development disorder," "PDD," "ASD"
Pediatric (in title or abstract)	"adolescent," "adolescence," "teen," "teenager," "young adult," "toddler," "preschool," "child," "children," "kindergarten," "middle school," "high school"
Recording (in any field)	"recordings," "recording," "tapes," "taped," "audio," "video," "measur*"

Note. NLS = natural language sampling; ASD = autism spectrum disorder; LENA = Language ENvironment Analysis.

leading these measures to be weakly correlated. More work is needed to identify gold-standard methods of extracting measures of language from recorded samples that are reliable and valid. This article will take steps toward this goal by identifying the most common measures of language derived from NLS and reviewing studies that have compared across NLS measures and against standardized measures.

Following an examination of the above questions, the review will conclude with a discussion of the remaining gaps in the literature, recommendations for employing NLS methods in research, and future directions for language research in autism that can be uniquely addressed using NLS.

Method

Search Strategy

To address the aims of this scoping review, a search of peer-reviewed empirical articles using NLS to measure language in samples of children diagnosed with autism was conducted following the guidelines set out in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (Tricco et al., 2018). A librarian conducted a literature search on 8 August 2023 in three databases: PsycInfo, Web of Science, and OVIDMedline. Search terms were selected to identify publications that included mentions of language, autism, children, and recordings (see Table 1). Search strategies and medical subject heading terms were modified for each database format. Searches were restricted to publications available in English. The search was restricted to works published in the past 15 years (after August 2008). After removing

duplicates, the search resulted in 4,671 publications, including one paper recommended by an expert.

Inclusion and Exclusion Criteria

Papers were assessed to determine if they met the following inclusion and exclusion criteria. First, only peer-reviewed empirical research papers including a sample of children with ASD were included. Therefore, papers with samples of participants that were over 18 years old were excluded. The criteria for ASD were broad; studies had to either assess for autism and specify their inclusion criteria or have a sample of participants with an existing self-reported diagnosis. Papers where NLS measures were collected prior to diagnosis (e.g., studies of infant siblings) and studies with samples younger than 18 months old were excluded from this review.

Studies had to include at least one natural language sample and at least one variable that was extracted from that sample. We defined a natural language sample as any collection of spontaneous language (i.e., sentence reading and book narration tasks were excluded) and involved interaction with a social partner who was also able to use spontaneous language. Studies that collected samples where the social partner was given scripted response were excluded. For example, Unwin et al. (2022) had an experimenter as the social partner who had a set of rote utterances they spoke in each condition and otherwise kept responses to the child to a minimum. Studies that measured samples where the social partner had some structured language but was also able to respond spontaneously were included.

One common method for collecting language samples was the Autism Diagnostic Observation Schedule (ADOS-2; Lord et al., 2012). The ADOS-2 is a diagnostic assessment that includes several different tasks that vary in how spontaneous, social, and flexible interactions are. For example, multiple tasks involve asking the child to create a narrative from provided materials such as a wordless story book or from cartoon images. Therefore, studies that only focused on the narration tasks or only reported findings from across the entire ADOS-2 were excluded from this review due to the limitations on spontaneity and social interaction. However, studies that selected only the conversation and questions (semistructured) sections of the ADOS-2 are included in the current review. Due to the focus on NLS as a measure of language, studies had to calculate at least one variable that could be collected from audio alone. Variables that *required* video coding were excluded (e.g., rates of social initiations that included vocalizations and gestures together). Studies could still use video to support coding, as long as one could in theory code the variable of interest from audio alone (e.g., a study may use video to support identifying who the speaker is when the variable of interest is rates of vocalizations).

Finally, case studies and reports on individual participants rather than group-level statistics were excluded. To make screening more efficient, any studies with 10 or fewer total participants were excluded. Additionally, papers that did not provide key methodological details (e.g., did not describe the NLS collection procedures) were excluded.

Screening Procedure

All titles and abstracts were reviewed by the first author; 4,441 abstracts were removed (see Figure 2 for reasons for exclusion). Due to the focus of this review on research methods, the Method section of the full texts were then reviewed by a team of six undergraduate research assistants. The full texts of the resulting 111 papers deemed possibly eligible for inclusion were then reviewed by the first author. A total of 59 publications met criteria for inclusion in the current review (Figure 2). All included publications are listed in Table 2 and denoted with a number that corresponds to Tables 3–5. Full citations can be found in the references section. Additional detailed results for included publications can be found in Appendix A in Supplemental materials.

Coding Procedure

Study characteristics for the variables of interest relevant to each aim were coded by the first author. Participant characteristics included age, sample groups, and language level. Collection methods that were coded included social partner, setting, context, length, and study goal. Both the methods of analyzing and types of variables were extracted to identify how measures of language were derived from NLS. Additional details on coding methods for the primary variables of interest are included below. Sex distribution in the ASD samples, sex of the caregiver, and the countries studies were conducted in were also noted when available. Studies that provided measures of reliability and validity were also identified.

Aim 1 Codes. Participant age was categorized into preschool-aged children (under 6 years old), Elementary school aged children (6–12 years old), and adolescents (12–18 years old). When participants span more than one age grouping, the range is reported.

Studies were coded for the inclusion of only autistic children, a typically developing comparison sample, and/or a nontypically developing comparison sample, such as children developmental delay but no ASD diagnosis.

Studies varied in the level of detail they reported about their participants' spoken language level at study entry. Participant language levels were categorized using criteria provided by Tager-Flusberg et al. (2009) and the ADOS-2 (Lord et al., 2012) descriptions of language levels, resulting in four categories: preverbal, single words, phrase speech,

and fluent speech. Preverbal describes children who are not yet producing words. Single words describes children who are not yet combining words; this stage occurs around 12 to 18-month-olds in typically developing children (Tager-Flusberg et al., 2009). Phrase Speech describes the stage when children combine words into short phrases, but are not yet speaking in full sentences. Finally, Fluent Speech describes children who are speaking in complete sentences and can have a conversation. In typical development, children begin to exit the phrase speech phase around 30–36 months, and their fluency continues to develop through the preschool years (Tager-Flusberg et al., 2009). When publications did not describe participants' language levels, details provided about ADOS modules administered and/or reports of expressive language age equivalents on standardized assessments were used to estimate language abilities (see Table 3). It is important to note that children are typically not classified as being nonspeaking or minimally speaking until after age 5. Additionally, studies of preschoolers typically did not have specific inclusion criteria related to language, but instead presented means and standard deviations of standardized assessment scores. This made it difficult to evaluate the full range of language in the sample. Therefore, we only describe the language level(s) of participants in studies that included children who were at least 6-years-old.

Aim 2 Codes. The social partner, setting in which the NLS was collected, context of the NLS, length of the NLS, and primary goal(s) of each study were identified. If studies had more than one type of natural language sample, we report the appropriate categories for each sample type. Social partner was coded as Peers (i.e., other children), Caregiver, Examiner (i.e., typically in an evaluation context), Interventionist or Teacher (i.e., a therapist, intervention provider, or classroom instructor), or Adult Confederate (i.e., an adult who is not a professional like an examiner and has some information about the NLS task). When there was no restriction on who the child could talk to, such as in daylong naturalistic recordings, the social partner was coded as Anyone.

Setting was coded as Lab when NLS were conducted in a research context, Clinic when NLS were collected in a treatment context, and Classroom when collected in a school environment (including therapeutic classrooms). Home-based collections were coded as Home, except when there were no restrictions on where participants could be, such as for daylong recordings. In these cases, we coded the setting as Anywhere.

We also identified the context of the language sample. Structured assessments, such as conversational portions of the ADOS-2, were coded as Assessment. We coded Classroom Interactions when participants interacted with peers and/or teachers in a classroom environment, and

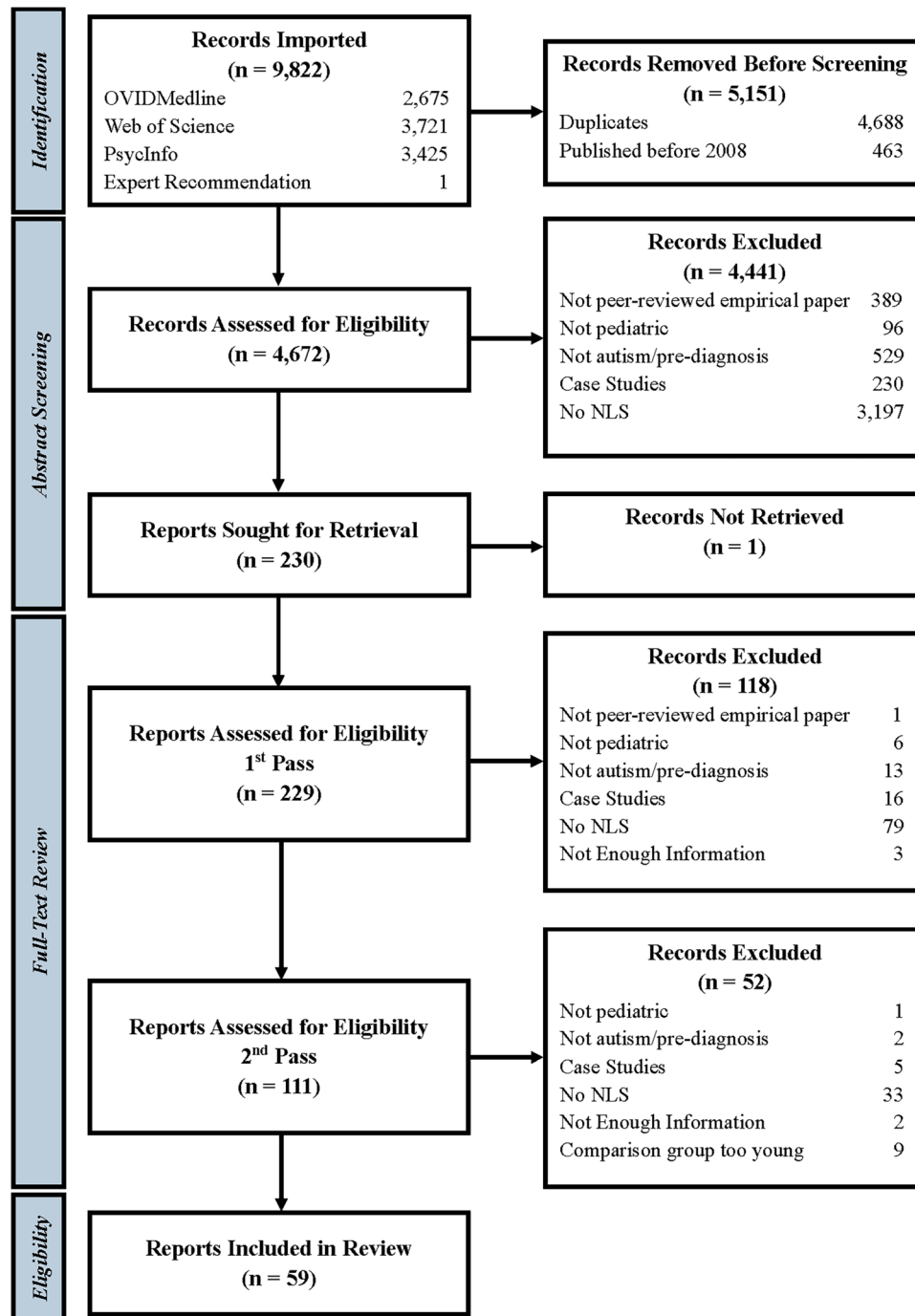


Figure 2. Publication Screening and Eligibility Process.

This PRISMA flow diagram displays the process for screening articles and assessing for eligibility and inclusion/exclusion criteria. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Intervention when the language sample was produced during an intervention or treatment session. Samples with no restrictions were coded as Everyday Life, whereas recording of everyday play at home were coded as Naturalistic Free Play. Play samples that included a standard set of toys, but not instructions for the social partner, were coded

as Play with Standard Toys. When the social partner was given instructions in addition to standard toys, or an established play task was used, we coded Structured Play. Finally, we coded context as Social Conversation when the task was entirely conversation based and did not fit any of the other contexts.

Table 2. Key for Included Publications.

Number	Citation (First author, year)
1	Adams et al., 2021
2	Alak, 2021
3	Alkire et al., 2023
4	Asgari et al., 2021
5	Barokova et al., 2020
6	Barokova et al., 2021
7	Bauminger-Zviely et al., 2020
8	Berman et al., 2018
9	Binns et al., 2022
10	Butler et al., 2022
11	Butler & Tager-Flusberg, 2022
12	Casenhiser et al., 2015
13	Cho et al., 2023
14	Clark-Whitney et al., 2022
15	Cola, Yankowitz, et al., 2022
16	Cola, Zampella, et al., 2022
17	De Korte et al., 2020
18	DiStefano et al., 2016
19	Elder et al., 2011
20	Fasano et al., 2023
21	Fasano et al., 2021
22	Fusaroli et al., 2019
23	Fusaroli et al., 2023
24	Germone et al., 2019
25	Goodwin et al., 2012
26	Gordon et al., 2011
27	Harbison et al., 2018
28	Hilvert et al., 2020
29	Kaale et al., 2018
30	Ko et al., 2019
31	Kover et al., 2014
32	Lai, 2020
33	Li et al., 2022
34	MacDonald-Prégent et al., 2023
35	MacFarlane et al., 2022
36	Maes et al., 2023
37	McDaniel & Brady, 2022
38	McDaniel et al., 2019
39	McDaniel et al., 2020
40	McDaniel et al., 2021
41	McKernan et al., 2022
42	McMahon et al., 2013
43	Nadig et al., 2010
44	Naigles et al., 2016
45	O'Haire et al., 2013
46	Parish-Morris et al., 2017
47	Pasco et al., 2008
48	Rollins et al., 2022
49	Song et al., 2021
50	Song & So, 2022
51	Thomas et al., 2022
52	Tovar et al., 2015
53	Trembath et al., 2019
54	Usher et al., 2015
55	Venker et al., 2015
56	Veytsman et al., 2023
57	Vivanti et al., 2019
58	Xu et al., 2021
59	Yoder et al., 2013

Natural language sample lengths were categorized as 10 min or less, greater than 10 min but less than 30, greater than 30 min but less than 3 hr, and Daylong. No studies had samples that were greater than 3 hr, unless they collected daylong recordings.

Finally, we identified the primary study goals related to NLS. Studies that primarily analyzed differences in NLS variables across groups were coded as Compare Groups. NLS measures often served as one component of a larger study looking at the relationship between language and other behaviors. In these cases, when NLS variables were the outcome measure we coded Dependent Variable and when these variables served as predictors, we coded Independent Variable. For studies using NLS to test the effects of an intervention we coded Intervention Effect. Lastly, some studies explicitly sought to validate the use of NLS as a measure of language. These studies were coded as NLS Validation. Studies could be coded for more than one primary goal.

Aim 3 Codes. To code the measures of language derived from NLS we identified both the methods of analysis and types of variables. Methods of analysis included Acoustic (i.e., extracting variables such as pitch from audio), Transcription (i.e., writing out words or phonemes), and Coding (i.e., behavioral coding that identifies properties of vocalizations). Additionally, if studies employed computerized, automated methods to extract variables from audio recordings, we coded Automated. Studies were coded for all methods used.

We identified up to three primary types of variables that were extracted from audio recordings. Variables categories included Acoustic (e.g., fundamental frequency), Vocalization Frequencies (e.g., rates of vocalizations), and Mean Length of Utterances. Consonant/Canonical was coded for variables such as consonant inventories or proportion of syllables that were canonical. Function captured studies that coded a purpose of verbal behaviors such as initiating or requesting. For studies that measured contingencies between child vocal behavior and the social partner, we coded Contingent. Studies that focused on the use of specific categories of words such as verbs or nouns were coded as Word Type. Finally, studies that focused on specific variables not captured in these categories were coded as Other.

Reliability. A second coder was trained on 12 files and provided feedback. Reliability coding was conducted on nine files (15% of the full sample) that were randomly selected. Percent agreement for each file was calculated and average agreement across files was 90.38% (min = 78.26%, max = 100%). All disagreements were discussed and consensus coded.

Results

Aim 1: Who is Included in NLS Research?

Characteristics of Participants. The results of Aim 1 are presented in Table 3. The samples of the 59 included studies included participants in the following age ranges: 25 (42.37%) preschool aged children (under 6 years old), five (8.47%) elementary years (6–12 years old), and one (1.69%) adolescents (12–18 years old). Many studies ($n = 28$, 47.46%) examined multiple age ranges, specifically 11 (18.64%) preschool to elementary aged children, 14 (23.73%) elementary children and adolescents, and three (5.08%) preschoolers to adolescents.

Samples also varied in size, from small samples of 13 to larger samples of up to 210. The average sample size across all studies was 63 participants ($M = 62.86$, $SD = 35.88$). Over half of studies ($n = 37$, 62.71%) included only samples of autistic children, 21 (35.59%) included a “typically developing” comparison group, and four (6.78%) included a nontypically developing comparison group (e.g., developmental delay, Williams Syndrome, Fragile X). Autistic participants were overwhelmingly male (mean percent male = 80.51, $SD = 10.85$, Range: 55.56%–100%).

Samples primarily consisted of children in the United States ($n = 44$, 74.58%), but NLS autism studies were also conducted in Norway ($n = 1$), Belgium ($n = 1$), Turkey ($n = 1$), Australia ($n = 2$), Israel ($n = 1$), Canada ($n = 3$),

China ($n = 3$), the Netherlands ($n = 1$), and the United Kingdom ($n = 2$).

Representation of Autistic Children with Differing Language Abilities. Table 3 describes the language levels of the 34 studies that included children who were at least 6-years-old. Of those studies, 14 also included preschool aged children, resulting in wide ranges in verbal abilities. Conversely, the 20 studies of children 6- to 18-years-old primarily included only children with fluent speech ($n = 14$, 70%). There is a clear dichotomy between studies of young children and studies of older children, such that children who do not have fluent speech by the time they are school-aged (i.e., are minimally or nonspeaking) are not represented in NLS research. Additionally, studies were highly variable in both their age and language ranges, which limits our understanding of how age, language ability, and use of NLS methods are related.

Aim 2: How are Natural Language Samples Collected?

NLS collection methods are summarized in Table 4. Language samples ranged in length from just a few minutes to day-long recordings. The majority ($n = 46$, 77.97%) of samples were shorter than 30 min. NLS measures were used for a variety of study goals, such as to compare groups

Table 3. Participant Characteristics for NLS Studies.

	<i>n</i> , %	Studies
Participant age	<i>n</i> = 59	
Preschool	25, 42.37%	2, 9, 12, 14, 21, 20, 22, 23, 25, 27, 29, 31, 36, 38, 39, 40, 44, 48, 51, 52, 53, 55, 57, 58, 59
Preschool-elementary	11, 18.64%	1, 10, 11, 17, 18, 19, 26, 33, 37, 47, 50
Elementary	5, 8.47%	4, 8, 31, 34, 43
Elementary-adolescence	14, 23.73%	3, 5, 7, 13, 15, 16, 28, 35, 42, 45, 46, 49, 54, 56
Adolescence	1, 1.69%	30
Preschool-adolescence	3, 5.08%	6, 24, 41
Sample composition	<i>n</i> = 59	
ASD only	37, 62.71%	2, 5, 6, 7, 8, 9, 10, 11, 12, 14, 17, 18, 19, 24, 26, 27, 29, 30, 31, 33, 34, 36, 37, 38, 39, 40, 41, 42, 47, 48, 50, 51, 52, 53, 55, 56, 57
Typical comparison	21, 35.59%	1, 3, 4, 13, 15, 16, 20, 21, 22, 23, 25, 32, 35, 43, 44, 45, 46, 49, 54, 58, 59
Nontypical comparison	4, 6.78%	20, 21, 28, 32
Language level*	<i>n</i> = 34	
Preverbal-single words	5, 14.71%	7, 18, 26, 37, 47
Preverbal-phrase speech	2, 5.88%	5, 6
Preverbal-fluent speech	3, 8.82%	10, 11, 34
Single words-fluent speech	3, 8.82%	17, 19, 33
Phrase speech-fluent speech	3, 8.82%	8, 28, 41
Fluent speech	16, 47.06%	1, 3, 4, 13, 15, 16, 30, 32, 35, 42, 43, 46, 49, 50, 54, 56
Not described	2, 5.88%	24, 45

Note. ASD = autism spectrum disorder; NLS = natural language sampling.

*Language level is not reported for studies that only included preschoolers.

Studies are listed by number which correspond with citations in Table 2.

Table 4. NLS Collection Methods.

	<i>n</i> , %	Studies
Social partner		
Adult confederate	6, 10.17%	13, 16, 43, 49, 56, 57
Anyone	2, 3.39%	27, 59
Caregiver	22, 37.29%	2, 5, 9, 10, 11, 12, 14, 17, 19, 22, 23, 25, 29, 31, 33, 36, 44, 48, 50, 52, 55, 58
Examiner	16, 27.12%	1, 4, 5, 6, 15, 28, 32, 35, 36, 37, 38, 39, 40, 41, 46, 51
Interventionist/teacher	10, 16.95%	8, 18, 20, 24, 26, 34, 42, 45, 47, 53
Peers	16, 27.12%	3, 7, 20, 21, 24, 26, 30, 41, 42, 45, 46, 47, 51, 53, 54, 57
Setting		
Anywhere	2, 3.39%	27, 59
Classroom	8, 13.56%	7, 20, 21, 26, 45, 47, 53, 57
Clinic	15, 25.42%	1, 3, 8, 12, 14, 17, 18, 24, 31, 34, 35, 36, 42, 50, 56
Home	14, 23.73%	5, 8, 10, 11, 14, 19, 22, 23, 25, 33, 36, 44, 48, 52
Lab	24, 40.68%	2, 4, 5, 6, 9, 13, 15, 16, 28, 29, 30, 32, 36, 37, 38, 40, 41, 43, 46, 49, 51, 54, 55, 58
Context		
Assessment	9, 15.25%	1, 4, 15, 32, 35, 37, 38, 46, 51
Classroom interactions	6, 10.17%	20, 21, 26, 47, 53, 57
Everyday life	2, 3.39%	27, 59
Naturalistic free play	10, 16.95%	2, 10, 11, 19, 22, 23, 25, 44, 52, 58
Intervention	6, 10.17%	8, 18, 24, 34, 42, 45
Play with standard toys	11, 18.64%	9, 12, 14, 18, 29, 31, 33, 36, 48, 50, 55
Social conversation	11, 18.64%	3, 7, 13, 16, 28, 30, 43, 49, 54, 56, 57
Structured play	12, 20.34%	5, 6, 17, 22, 23, 25, 39, 40, 41, 44, 52, 58
Length		
Less than 10 min	15, 25.42%	3, 7, 8, 13, 16, 24, 30, 32, 35, 40, 43, 45, 49, 54, 56
10–30 min	31, 52.54%	1, 2, 4, 5, 6, 9, 10, 11, 12, 14, 15, 18, 19, 26, 28, 29, 31, 34, 36, 37, 38, 39, 41, 42, 46, 48, 50, 51, 55, 57, 58
30 min–3 hr	10, 16.95%	17, 22, 23, 25, 33, 44, 47, 52, 53, 57
Daylong	4, 6.78%	20, 21, 27, 59
Study goal		
Compare groups	21, 35.59%	3, 5, 11, 13, 15, 16, 17, 19, 20, 21, 22, 25, 28, 32, 43, 44, 45, 46, 49, 54, 58
Dependent variable	10, 16.95%	2, 9, 10, 11, 23, 29, 33, 44, 50, 55
Independent variable	6, 10.17%	20, 25, 31, 35, 51, 52
Intervention effect	16, 28.81%	7, 8, 12, 14, 18, 19, 24, 26, 30, 34, 41, 42, 48, 53, 56, 57
NLS validation	12, 20.34%	1, 4, 6, 27, 35, 37, 38, 39, 40, 41, 47, 59

Note. NLS = natural language sampling.

Studies are listed by number which correspond with citations in Table 2.

($n = 21$, 35.59%), measure intervention effects ($n = 17$, 28.81%), or study the relations between language and other behaviors ($n = 16$, 27.12%). Additionally, 12 studies (20.34%) explicitly aimed to validate NLS measures. For example, one study validated measures of vocal communication (number of communication acts with a vocalization, proportion of vocalizations that are communicative) and vocal complexity (diversity of key consonants in communication acts, proportion of vocalizations with a canonical syllable; McDaniel et al., 2020). The authors found that these measures had convergent validity, divergent validity, and sensitivity to change in a sample of preschoolers during assessment and free play across a year (McDaniel et al., 2020).

Other studies identified the length and/or number of recordings needed to create a stable measure of language. For a more complex measure—vocal reciprocity (likelihood of a child responding to an adult)—stability was acquired from two daylong recordings and vocal reciprocity was related to other measures of language, such as how “speech-like” the child’s vocalizations were (Harbison et al., 2018). McDaniel and Brady (2022) found that different vocal measures required different sample lengths to reach reliable measurement, and that sample length impacted convergent validity. Broadly, 10-min samples were sufficient for establishing reliability and validity across all vocal measures in a sample of children who did not yet have phrase speech (McDaniel & Brady, 2022).

Setting and Context. Participants interacted with a variety of social partners, including caregivers, peers, interventionists, examiners, and novel adult confederates. Most studies included an adult social partner ($n = 54$, 91.53%). Additionally, most social partners were assumed to be neurotypical; studies generally did not specify whether the partner had a diagnosis of autism or other neurodevelopmental disorder. Half of the studies examining interactions with peers included autistic children as the social partners ($n = 8$, 50%), many of whom were other students in an intervention classroom. For example, one study used LENA recording devices to capture children's language in an autism intervention classroom where possible social partners were peers and teachers (Trembath et al., 2019). Another study more explicitly looked at how autistic children interact with autistic versus nonautistic peers by having ASD–ASD, non-ASD–non-ASD, and ASD–non-ASD dyads engage in get-to-know-you conversations (Alkire et al., 2023). This study used unfamiliar social partners, but just over half of studies employed familiar social partners ($n = 33$, 55.93%). Finally, of the 22 studies looking at caregiver–child interactions, 10 (45.45%) did not specify who the caregiver was and eight included primarily mothers (only mothers, $n = 5$; over 80% mothers, $n = 3$).

Samples were primarily collected in a research laboratory ($n = 24$, 40.68%). Other settings included participants' homes ($n = 14$, 23.73%), clinical settings (15, 25.42%), and classrooms ($n = 8$, 13.56%). Two studies using daylong recordings allowed participants to go about their daily lives, and therefore the setting was not standardized across children.

In addition to considering the setting, it is important to consider the context in which NLS measures are collected. Allowing children to play freely, either with their own toys ($n = 10$; 16.95%) or a standard set of toys ($n = 11$; 18.64%), was common. More structured play contexts were also employed ($n = 12$; 20.34%), such as the Eliciting Language Samples for Analysis (Barokova et al., 2021) or an adaptation of the Screening Tool for Autism in Two-Year-Olds (Stone et al., 2004), administered by parents. For older children, social conversations were a common context ($n = 11$; 18.64%). However, other studies used less naturalistic contexts, capturing language samples during assessments ($n = 9$; 15.25%) or interventions ($n = 6$; 10.17%). Intervention contexts varied and included playing with peers with and without an animal present (Germone et al., 2019; O'Haire et al., 2013), playing with an interventionist and caregiver during Pivotal Response Treatment (Berman et al., 2018), game time with peers during a social skills group-based intervention (McMahon et al., 2013), interactions with an interventionist during communication intervention with and without a speech-generating device (DiStefano et al., 2016), and music-mediated and play-based interventions for language (MacDonald-Prégent et al., 2023).

How Naturalistic are NLS Methods? Examining social partners developmentally, preschoolers primarily interacted with familiar adults, such as caregivers, teachers, and interventionists. Only six of the 39 studies that included children under 6-years-old observed interactions with peers. Studies of older children were more likely to include interactions with unfamiliar adults (e.g., examiners, confederates) and nine of the 34 studies that included children over age 6 sampled interactions with peers. These included peers who were autistic, nonautistic, familiar, and novel. Overall, social partners were highly variable, and while there were trends in the utilization of different social partners by age, many studies had children with a wide range of ages interact with the same type of social partner, which may not be developmentally appropriate.

Aim 3: How are Measures of Language Derived from NLS?

After collecting recordings of natural language, samples must be coded to extract measures of language. Table 5 presents both the methods used to analyze samples and the types of variables that were extracted. NLS recordings were primarily transcribed ($n = 34$, 57.63%) and/or behaviorally coded ($n = 34$, 57.63%). Additionally, many transcription studies specified that they used standardized formats and guidelines such as Computerized Language Analysis (CLAN)/Codes for the Human Analysis of Transcripts (CHAT; MacWhinney, 2000) or Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2012) guidelines. Behavioral coding designs varied and were often study specific. Few studies made use of acoustic analyses ($n = 3$; 5.08%). Several studies used automated methods such as LENA ($n = 7$; 11.86%), only one of which included participants older than preschool age (McKernan et al., 2022). Of these studies, two examined the reliability of automated coding with human coding (Fasano et al., 2021, 2023), which was high.

One of the many advantages of NLS is that researchers can extract a large variety of variables from recordings. These include frequencies of verbalizations (e.g., words, vocalizations, babbles; $n = 44$, 74.58%), functions of verbal behaviors (e.g., initiate, request; $n = 13$, 22.03%), consonant inventories or the use of canonical syllables ($n = 5$, 8.47%), and the use of different types of words (e.g., verbs, pronouns; $n = 9$, 15.25%). Several studies also calculated a common measure of language ability from transcripts—mean length of utterance (MLU), or the average number of units of speech in an utterance ($n = 19$, 32.20%). Because NLS vocalizations are produced in the context of a social interaction, studies also evaluated the contingencies between the verbal behavior of social partners ($n = 7$, 11.86%). Multiple measures from NLS were often used to address study aims, and while there were some measurements that were common (e.g., MLU, frequency of

Table 5. Measures Derived from NLS.

	<i>n, %</i>	Studies
Method of analyzing		
Acoustic	3, 5.08%	4, 13, 35
Automated	7, 11.86%	20, 21, 27, 41, 53, 57, 59
Coding	34, 57.63%	2, 3, 6, 7, 8, 12, 14, 17, 18, 19, 24, 25, 26, 28, 29, 30, 31, 32, 33, 36, 37, 38, 39, 40, 42, 43, 44, 45, 47, 48, 54, 56, 57, 58
Transcription	34, 57.63%	1, 3, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 18, 22, 23, 25, 28, 31, 32, 33, 34, 35, 36, 39, 43, 44, 46, 48, 49, 50, 51, 52, 55, 58
Variables		
Acoustic	2, 3.39%	4, 13
Contingent	7, 11.86%	1, 6, 13, 18, 27, 30, 41
Consonant/canonical	5, 8.47%	36, 37, 38, 39, 40
Function	13, 22.03%	8, 12, 17, 18, 26, 30, 31, 33, 43, 47, 48, 54, 56
Frequencies	44, 74.58%	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 19, 20, 21, 22, 23, 24, 26, 28, 29, 31, 34, 36, 37, 38, 39, 40, 42, 43, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 57
MLU	19, 32.20%	1, 5, 10, 11, 12, 14, 22, 23, 25, 28, 31, 33, 34, 43, 44, 50, 51, 52, 55
Word types	9, 15.25%	8, 15, 25, 32, 44, 46, 49, 52, 58
Other	3, 5.08%	28, 35, 59

Note. NLS = natural language sampling; MLU = mean length of utterance. Studies are listed by number which correspond with citations in Table 2.

utterances, number of different words), studies also devised unique measures to answer their questions of interest. These include the ratio of “um” to “uh” utterances (Parish-Morris et al., 2017), number of “wh-” questions (Goodwin et al., 2012), and rating of topic maintenance (Nadig et al., 2010).

Are NLS Measures Reliable and Valid? One important question when considering NLS as a measure of language in autism research is how reliable and valid these methods are. When conducting behavioral coding or transcription of audio, calculation and maintenance of intercoder reliability is critical. However, examination of the included studies revealed that there is no standard way to conduct interrater reliability during training and data coding. Another concern is that many studies reported very few details about reliability, often including vague statements such as, “coders attained research reliability.”

Some studies assessed the stability of NLS over time (see McDaniel et al., 2021; Yoder et al., 2013), suggesting that as little as one observation is needed to achieve stable measures of language. Others evaluated concurrent, construct, and divergent validity, showing that NLS measures are related to concurrent standardized measures of language (Barokova et al., 2021; Harbison et al., 2018; McDaniel et al., 2020; Yoder et al., 2013). For example, Barokova et al. (2021) found that frequency of utterances and conversational turns during structured play were correlated with caregiver report measures of communication more broadly and had high test–retest reliability. Similarly, variables such as diversity of consonants and frequency of vocalizations were found to predict composite expressive language scores in preschoolers 12 months later (McDaniel et al., 2020).

One study also provided evidence that short samples (under 10 min) are valid measures of language (McDaniel & Brady, 2022). These studies provide preliminary evidence that NLS can be collected briefly and are a valid granular measure of language development.

Discussion

With the rise of NLS methods in autism language research, it is essential to review the current state of the field to identify common practices and future directions for advancement of the understanding of language in autism. This scoping review adds to the conversation about using NLS methods in autism language research by discussing the advantages and disadvantages of common methods in the context of development and what it means to be “naturalistic.” This review identified 59 studies that used NLS methods to measure expressive language in autistic children over the past 15 years. These publications examined preschoolers to adolescents, individuals with preverbal to fluent speech, and sampled language for 3 min to multiple days using a wide variety of methods. The discussion below highlights how underreporting the details of study methods complicates comparing and drawing conclusions across studies. Despite the variability, several key patterns of common study designs emerged and are discussed below.

Who is Included in NLS Research?

In examining the publications in the current review, a dichotomy emerged, where studies tended to sample *either* younger children who have preverbal to phrase speech

language abilities or older children and adolescents who use fluent speech. School-age children and adolescents who do not have fluent speech (i.e., minimally or nonspeaking) were underrepresented. Although NLS is uniquely able to capture language skills and variability that traditional standardized measures of language cannot, the advantages of this method have not been realized across all age and ability groups.

In addition, many publications included participants with large age ranges. When studying skills that change with development, such as expressive language, it is important to consider age and developmental level. Publications measuring language in children with autism face a particular challenge of highly heterogeneous samples. While some studies included comparison groups (37%), only three studies included both an age-matched and a language-matched sample in the same study (Fasano et al., 2021, 2023; Lai, 2020).

Samples were primarily male, which is in line with current estimates that 3.8 boys are diagnosed autistic for every one girl (Maenner, 2023). However, the decreased prevalence of autism in girls compared to boys is reinforced by the male-centered conceptualization of autism, which is further perpetuated by research findings that are based on primarily male samples. Additionally, no studies reported including gender-diverse individuals, even in samples of adolescents. Finally, the majority of studies were conducted in the United States, and all other countries in this review were represented by three or fewer studies.

Based on these findings, future studies should consider the following recommendations. First, more research is needed that includes autistic females and examines gender and sex differences in language in autistic youth. Additionally, there is a large gap in the use of NLS cross-culturally. Given that components of language, particularly pragmatics, differ by culture, it is essential that future work capitalize on NLS to study language in diverse samples across the world. Studies should also focus on narrower age ranges and language ability ranges. Sampling language from 6- and 16-year-olds using the same NLS method may not be developmentally appropriate because *natural* interaction contexts differ for young children and adolescents. Additionally, these large age ranges make it difficult to draw conclusions about the development of expressive language skills in autism. Similarly, many studies included children who communicate with single words in samples with children who have fluent speech. Therefore, researchers should consider focusing their samples on a particular developmental stage or language level, or report results both on the full sample *and* by age or language level (see Barokova et al., 2021 and Butler & Tager-Flusberg, 2023 for examples).

It is also recommended that when comparison samples are included, information is provided about whether the sample is matched by age and/or language level. When

possible, study designs should include both age-matched and language-matched comparison groups. Language-matched samples are useful for separating out differences that may be specific to autism, rather than driven by differences in expressive language development. Samples of children with other diagnoses, such as developmental delay or Down syndrome, may be appropriate depending on the study goal.

Publications should also report more explicitly on the language abilities of their sample at study entry. For example, some studies reported expressive language age equivalents from a standardized assessment but did not include the range for the sample. Similarly, many studies included ADOS-2 administration in their study procedures, but not all studies reported which modules were administered. Very few studies explicitly stated language level inclusion/exclusion criteria for participants. It is recommended that studies provide more detailed descriptions of the participants' language levels, either by broad expressive level (e.g., preverbal, single words, phrase speech, and fluent speech) or provide a measure of language developmental age and report minimum and maximum observed values. These descriptions are important because studies that examine language abilities in children with fluent speech are likely to have very different findings compared to studies of children who are preverbal or use single words. Failing to provide these details makes it difficult for readers to contextualize the results of the study. Additionally, studies of older autistic children might be assumed to be reporting on children with fluent speech when information about language ability is lacking, leading to the continued lack of representation of autistic children with minimal or no spoken language.

Finally, studying language in older children with minimal or no speech using NLS should be a research priority. This review supports previous calls for using NLS in this population (Kasari et al., 2013; Tager-Flusberg & Kasari, 2013) and highlights the significant lack of research in this area. In line with evidence that NLS is one of the most appropriate methods for measuring language in autistic children with minimal or no spoken language (Kasari et al., 2013), future studies with this population should include a brief naturalistic language sample in their study design. NLS methods can be used to identify components of language, such as phonology (e.g., consonant inventories) and pragmatics (e.g., contingent vocalizations). Additionally, researchers who already employ NLS methods for school-age children and adolescents should consider expanding their research to include autistic individuals who do not have fluent speech (Tager-Flusberg & Kasari, 2013).

How are Natural Language Samples Collected?

A primary goal of this review was to define “naturalistic” language samples by identifying the most common methods

and considering how naturalistic they are. To this end, included studies were examined to identify the settings and contexts in which, and the social partners with whom NLS were collected. While there was variability in methods, broadly speaking NLS were collected during everyday unstructured interactions (e.g., classroom interactions, interventions, and everyday life) or in play-based interactions (e.g., free play with or without standard toys). For older children, social conversations were also a common context. However, some studies had more structure to their data collection, either using structured play to elicit language, or using language collected during assessments. While collecting language during assessments (e.g., ADOS-2) is convenient, the context is not one that children experience in their daily lives. Further, the social partner has limitations on how naturally they can respond. Given that the majority of NLS are less than 30 min, and that several studies collected samples of just 10 min or less, capturing language using NLS in addition to typical standardized assessments could strengthen study designs with limited additional burden on participants and researchers.

Social Partners. While it is useful to understand how children communicate with novel communication partners, researchers must consider their research question and whether a more naturalistic sample that reflects a child's typical daily interaction partners should be prioritized. Children primarily interacted with adults in the studies reviewed here, and these adults were usually familiar to the child (e.g., caregiver, interventionist, and teacher). Some studies also included peers both with autism (e.g., other participants in an intervention) and without autism (e.g., other students in an inclusion classroom). Only two studies (Ko et al., 2019; Usher et al., 2015) included novel peers to observe how children got to know unfamiliar children.

As few studies focused on different types of peers, future research is needed to understand how autistic children interact with other autistic children versus nonautistic children, and novel versus familiar peers. This work is particularly needed given recent literature describing the double empathy problem in autistic adults. The double empathy problem posits that challenges with communication, reciprocity, and rapport between autistic and nonautistic individuals is due to mutual difficulty in understanding the other person, rather than from deficits specific to the autistic person (Davis & Crompton, 2021; Milton, 2012). Indeed, research with adults has found that autistic–nonautistic dyads have lower rapport, less effective communication, and less self-disclosure than autistic–autistic dyads and nonautistic–nonautistic dyads (Crompton et al., 2020; Morrison et al., 2020; Rifai et al., 2022). Despite clear differences in how same and different neurotype pairs interact, there is little work to date examining social communication and the

double empathy problem developmentally (Davis & Crompton, 2021).

Setting and Context. NLS were primarily collected in lab settings, but leaving the lab and capturing language in the community is important to validate whether lab-based data reflect a child's language ability. In fact, one study (Barokova et al., 2020) showed that parents elicited *more* language from minimally speaking children at home than experimenters did in the lab. This difference calls into question the generalizability of lab-based studies, particularly for children who may be more sensitive to novel environments and people. Future work is needed to compare how autistic children use language in settings of various levels of familiarity.

Notably, more than a quarter of studies (28.81%) used NLS to measure intervention effects. NLS provides several advantages for measuring change in language compared to standardized measures. Interventions are often short-term, requiring outcome measures to identify changes over a few months. Standardized measures are designed to be stable over time and are often not sensitive to change. Additionally, changes in social communication in response to intervention may be nuanced, small, and gradual. For example, a child who receives an intervention to improve expressive language may only learn to say five more words measured on an expressive language scale. However, that same child may increase the number of vocalizations they direct to a caregiver, the proportion of vocalizations that includes word approximations, and use the words they do know more frequently. In this case, NLS could capture these changes during a brief caregiver–child interaction. Additionally, behaviors produced in these samples can be understood in their relationship to everyday behavior, while standardized assessments do not have clear connections to what children do in their everyday interactions.

NLS as a Developmentally Appropriate Measure of Language.

As noted above, children's primary social partners change with development, as do the contexts in which they typically interact with others. If researchers want to capture truly naturalistic language, they must record children in their natural environments and with familiar interlocutors. One recommendation is that researchers take an ecological systems approach when designing their studies (Bronfenbrenner, 1977) and consider the many environments and people in the child's microsystem.

First, while caregivers are important social partners, especially in the early years, understanding how children and adolescents communicate with caregivers is unlikely to inform how they use language to interact with novel adults, familiar peers, and new romantic partners. Thus, we must consider including other people in the child's everyday environment. For example, once children begin school, they spend a large portion of their time interacting

with other students. Given the focus on peer interactions in the clinical diagnosis of autism, opportunities for capturing communication in the school environment are particularly salient. And for adolescents, this may mean capturing how they communicate with romantic partners, at community-based activities, and at first jobs. What is most “natural” will vary by the sample of participants one is interested in studying.

Second, if we want to know about the social communication challenges autistic children face in everyday life, we need to observe them in those contexts. While research settings provide more control and consistency when collecting samples, children may not behave as they typically would due to the unfamiliar environment. To capture truly “natural” language, we must sample children in their typical environments—at home, at school, on the playground, and at extracurricular activities. Similarly, while assessment and intervention contexts are convenient for capturing language given that they are often already recorded, researchers must consider the limitations of these contexts. Assessments are not necessarily representative of how children behave in their everyday lives (Tamis-LeMonda et al., 2017). It is recommended that researchers consider how the context used to sample language may, or may not be, natural for the child.

How are Measures of Language Derived from NLS?

While one advantage of NLS is that a very large number of measures can be extracted to meet the goals of the study, a disadvantage is that studies calculate and report very different measures. To compare language across studies, publications must report similar measures. This review identified the most common measures extracted from NLS, including frequencies, mean length of utterance, and function of vocalizations. Most studies used transcription and/or behavioral coding completed by human coders. Transcription studies tended to use standardized and widely available procedures, whereas behavioral coding procedures varied based on study design and goal. Additionally, there was variability in how much detail authors provided about coding and reliability procedures, as well as coder training. This variability makes it difficult for other researchers to replicate methods across new samples and contexts. Therefore, it is recommended that researchers provide details about coder training, coding manuals, and information regarding calculation of primary measures either in the main publication text or in supplemental materials. Additionally, CHAT/CLAN (MacWhinney, 2000) and SALT (Miller & Iglesias, 2012) are validated and widely used methods of child language transcription. Future transcription studies should utilize these methods and describe any deviations from these procedures.

Regardless of the primary variables of interest, there are a few language variables that all studies should report when available. Based on the current prevalence in the literature,

it is recommended that, when possible, studies report mean length of utterance (in morphemes or words), rate/frequency of utterances, and number of different words (or consonants if studying children who do not use words). Providing these descriptive data will allow for future work to compare findings from multiple studies.

Finally, 20% of publications explicitly aimed to validate NLS measures. However, there was not consistency in the variables that were studied, nor in the methods used. Future work should continue to validate NLS measures in different samples and contexts.

Future Directions

This review has highlighted the many advantages of using NLS to study language in autism and the many gaps in the current literature. There are endless novel questions that NLS is well suited to help answer. To address these open questions, research must continue to move out of the lab and into children’s environments to ask about their language-related strengths and challenges in everyday interactions. How do autistic children communicate with unfamiliar peers on the playground? How do they resolve conflicts with friends? How do adolescents interact with coworkers or get to know potential romantic partners? Exploring questions like these, which focus on how autistic children and adolescents use language to navigate everyday interactions, will identify challenges and provide new directions for developing potential supports. Additional work should also be conducted to ask autistic individuals and their families about their research priorities with regards to studying language in autism.

Additionally, while studies of infants have made extensive use of acoustic measures to identify potential vocal markers that portend later diagnosis, NLS research in autistic children has made minimal use of these measures. For example, the infant literature has made use of the automated coding provided by LENA to identify adult and child vocalizations and validated this method for young infants later diagnosed with autism (Dykstra et al., 2013; Wang et al., 2017). However, there has been limited work validating similar systems for older autistic children, and only one study in this review employed automated methods in children over 6 years of age. This dearth of research using automated methods with older children suggests that there is not yet an automated measure that can be reliably used to measure expressive language characteristics in autistic children. Availability of such measures and tools for language coding would increase the feasibility of collecting longer samples from larger numbers of children. The rise of natural language processing and machine learning models in other fields provides exciting new avenues for NLS research in autism. To take advantage of these cutting-edge methods however, we must validate them with large datasets of language samples from well

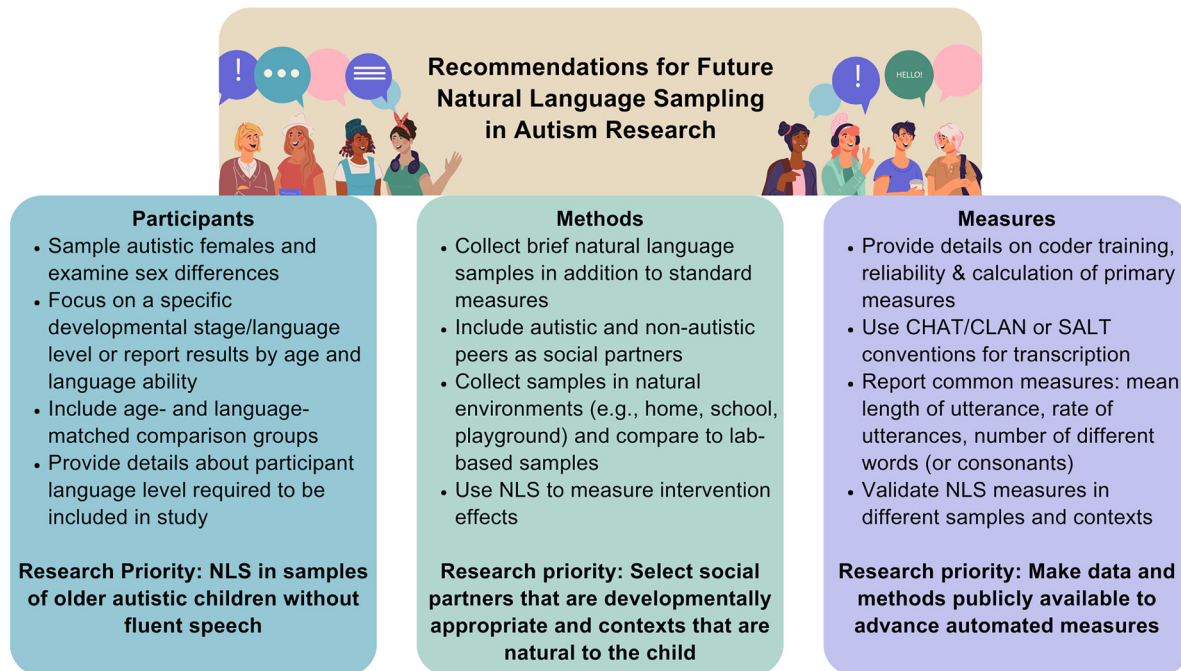


Figure 3. Recommendations for Future Research.

phenotyped autistic and nonautistic children. One recommendation is that studies make their data and methods publicly available through research repositories such as Databrary (databrary.org) and Child Language Data Exchange System (childes.talkbank.org). Additionally, researchers who develop well-validated language models should make those methods publicly available so that they can begin to be used with other datasets. Improving automated methods would allow for more naturalistic studies in which autistic children are provided with a small recording device and both their language and the language environment around them is captured for many hours across many days.

Finally, NLS is a highly granular and sensitive measure and should be used in studies where standardized measures, which have floor effects and are insensitive to changes over time, may not be appropriate. As is evident in this review, intervention studies targeting language have already begun to use NLS as a primary measure of intervention effects. One large gap in the literature, however, has been the use of NLS with older autistic children who produce minimal spoken language. NLS can provide additional insight in these children's communication, above and beyond standardized language, by quantifying measures of developing speech, such as rates of vocalizations, frequencies of prespeech sounds like canonical syllables, and patterns of initiation with, and contingent responses to, social partners. A clinical advantage of natural observation is that it can provide clinicians with information about exactly where a child's skills are so they can promote increased use of new skills during intervention.

Conclusions

While work across the past 15 years has made many advancements in using NLS to measure expressive language in autistic children, including validation of specific measures and using NLS to evaluate response to intervention, there has been limited consensus around the best methods and measures. The current review outlined recommendations for researchers to consider when selecting participant groups, sampling methods, and measures for NLS studies (Figure 3). This review was limited in scope, specifically with regards to characterizing the methods used but not evaluating the quality of the studies or results. Additional work is needed to address the gaps described here and replicate previous findings to identify patterns of natural language across the literature. In particular, there is an urgent need to improve our understanding of autistic children with minimal or no spoken language and high support needs. Funding agencies have also taken notice of this gap in the literature, as can be seen in the 2023 National Institutes of Health Notice of Special Interest: Promoting Language and Communication in Minimally Verbal/Non-Speaking Individuals with Autism (NOT-DC-23-009). It is vital that we improve NLS methods for this population to increase access to research studies, as a deeper understanding of the language and communication challenges in this population can lead to clinical advancements to support those most affected by ASD.

Acknowledgments

I would like to thank Dr Jana Iverson for her immensely helpful feedback on this manuscript. Her input and thoughtful consideration are invaluable. Many thanks to Drs Daniel Shaw, Jessie Northrup, Sophia Choukas-Bradley, and Joshua Schneider for their comments on a previous version of this manuscript; their suggestions have greatly improved this review. I am also grateful to the members of the Infant Communication Lab at the University of Pittsburgh and Boston University for their support. Portions of this work were presented at the Meeting on Language in Autism 2024.

ORCID iD

Samantha Nicole Plate  <https://orcid.org/0000-0001-8812-5534>

Funding

During preparation of this article, SNP was supported by a grant awarded to Jana Iverson from the National Institute on Deafness and other Communication Disorders (Grant No. R01 DC016557).

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

All data generated or analyzed during this study are included in this published article and its supplemental information files.

Note. Reviewed papers are denoted with a number

Supplemental material

Supplemental material for this article is available online.

References

- Adams, J. R., Salem, A. C., MacFarlane, H., Ingham, R., Bedrick, S. D., Fombonne, E., Dolata, J. K., Hill, A. P., & Van Santen, J. (2021). A pseudo-value approach to analyze the semantic similarity of the speech of children with and without autism spectrum disorder. *Frontiers in Psychology, 12*, 668344. <https://doi.org/10.3389/fpsyg.2021.668344>
- Alak, G. (2021). Otizm spektrum bozukluğu olan Çocuk anneleri Çocuklarının hangi davranışlarını nasıl yaşıyor? *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi, 22*(2), 465–489. <https://doi.org/10.21565/ozelegitimdergisi.699841>
- Alkire, D., McNaughton, K. A., Yarger, H. A., Shariq, D., & Redcay, E. (2023). Theory of mind in naturalistic conversations between autistic and typically developing children and adolescents. *Autism, 27*(2), 472–488. <https://doi.org/10.1177/136236132211036>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American psychiatric association. <https://doi.org/10.1176/appi.books.9780890425596>
- Arunachalam, S., Avtushka, V., Luyster, R. J., & Guthrie, W. (2022). Consistency and inconsistency in caregiver reporting of vocabulary. *Language Learning and Development, 18*(1), 81–96. <https://doi.org/10.1080/15475441.2021.1931233>
- Asgari, M., Chen, L., & Fombonne, E. (2021). Quantifying voice characteristics for detecting autism. *Frontiers in Psychology, 12*, Article 665096. <https://doi.org/10.3389/fpsyg.2021.665096>
- Barokova, M., Hassan, S., Lee, C., Xu, M., & Tager-Flusberg, H. (2020). A comparison of natural language samples collected from minimally and low-verbal children and adolescents with autism by parents and examiners. *Journal of Speech, Language, and Hearing Research, 63*(12), 4018–4028. https://doi.org/10.1044/2020_JSLHR-20-00343
- Barokova, M., La Valle, C., Hassan, S., Lee, C., Xu, M., McKechnie, R., Johnston, E., Krol, M. A., Leano, J., & Tager-Flusberg, H. (2021). Eliciting language samples for analysis (ELSA): A new protocol for assessing expressive language and communication in autism. *Autism Research, 14*(7), 1529–1529. <https://doi.org/10.1002/aur.2544>
- Barokova, M., & Tager-Flusberg, H. (2018). Commentary: Measuring language change through natural language samples. *Journal of Autism and Developmental Disorders, 50*(7), 2287–2306. <https://doi.org/10.1007/s10803-018-3628-4>
- Bauminger-Zviely, N., Estrugo, Y., Samuel-Magal, K., Friedlin, A., Heishrik, L., Koren, D., & Bar-Yehuda, S. (2020). Communicating without words: School-based RCT social intervention in minimally verbal peer dyads with ASD. *Journal of Clinical Child & Adolescent Psychology, 49*(6), 837–853. <https://doi.org/10.1080/15374416.2019.1660985>
- Berman, S., Ventola, P., & Gordon, I. (2018). Improvements in micro level indices of social communication following pivotal response treatment (PRT). *Research in Autism Spectrum Disorders, 51*, 56–65. <https://doi.org/10.1016/j.rasd.2018.04.003>
- Binns, A. V., Casenhiser, D. M., Shanker, S. G., & Cardy, J. O. (2022). Autistic preschoolers' engagement and language use in gross motor versus symbolic play settings. *Autism & Developmental Language Impairments, 7*, Article 239694152211150. <https://doi.org/10.1177/23969415221115045>
- Boorse, J., Cola, M., Plate, S., Yankowitz, L., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2019). Linguistic markers of autism in girls: Evidence of a “blended phenotype” during storytelling. *Molecular Autism, 10*(1), Article 14. <https://doi.org/10.1186/s13229-019-0268-2>
- Brady, N. C., Romine, R. E. S., Holbrook, A., Fleming, K. K., & Kasari, C. (2020). Measuring change in the communication skills of children with autism spectrum disorder using the communication complexity scale. *American Journal on Intellectual and Developmental Disabilities, 125*(6), 481–492. <https://doi.org/10.1352/1944-7558-125.6.481>
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*(7), 513–531. <https://doi.org/10.1037/0003-066X.32.7.513>
- Butler, L. K., La Valle, C., Schwartz, S., Palana, J. B., Liu, C., Peterman, N., Shen, L., & Tager-Flusberg, H. (2022).

- Remote natural language sampling of parents and children with autism spectrum disorder: Role of activity and language level. *Frontiers in Communication*, 7, Article 820564. <https://doi.org/10.3389/fcomm.2022.820564>
- Butler, L. K., & Tager-Flusberg, H. (2023). Fine motor skill and expressive language in minimally verbal and verbal school-aged autistic children. *Autism Research*, 16(3), 630–641. <https://doi.org/10.1002/aur.2883>
- 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. <https://doi.org/10.1007/s10803-014-2242-3>
- Chen, Y., Siles, B., & Tager-Flusberg, H. (2024). Receptive language and receptive-expressive discrepancy in minimally verbal autistic children and adolescents. *Autism Research*, 17(2), 381–394. <https://doi.org/10.1002/aur.3079>
- Cho, S., Cola, M., Knox, A., Pelella, M. R., Russell, A., Hauptmann, A., Covello, M., Cieri, C., Liberman, M., Schultz, R. T., & Parish-Morris, J. (2023). Sex differences in the temporal dynamics of autistic children's natural conversations. *Molecular Autism*, 14(1), Article 13. <https://doi.org/10.1186/s13229-023-00545-6>
- Clark-Whitney, E., Klein, C. B., Hadley, P. A., Lord, C., & Kim, S. H. (2022). Caregiver language input supports sentence diversity in young children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 65(4), 1465–1477. https://doi.org/10.1044/2021_JSLHR-21-00458
- Cola, M., Yankowitz, L. D., Tena, K., Russell, A., Bateman, L., Knox, A., Plate, S., Cubit, L. S., Zampella, C. J., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2022). Friend matters: Sex differences in social language during autism diagnostic interviews. *Molecular Autism*, 13(1), Article 5. <https://doi.org/10.1186/s13229-021-00483-1>
- Cola, M., Zampella, C. J., Yankowitz, L. D., Plate, S., Petrulla, V., Tena, K., Russell, A., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2022). Conversational adaptation in children and teens with autism: Differences in talkativeness across contexts. *Autism Research*, 15(6), 1090–1108. <https://doi.org/10.1002/aur.2693>
- Crompton, C. J., Sharp, M., Axbey, H., Fletcher-Watson, S., Flynn, E. G., & Ropar, D. (2020). Neurotype-matching, but not being autistic, influences self and observer ratings of interpersonal rapport. *Frontiers in Psychology*, 11, Article 586171. <https://doi.org/10.3389/fpsyg.2020.586171>
- Davis, R., & Crompton, C. J. (2021). What do new findings about social interaction in autistic adults mean for neurodevelopmental research? *Perspectives on Psychological Science*, 16(3), 649–653. <https://doi.org/10.1177/1745691620958010>
- De Korte, M. W., Van Den Berk-Smeekens, I., Van Dongen-Boomsma, M., Oosterling, I. J., Den Boer, J. C., Barakova, E. I., Lourens, T., Buitelaar, J. K., Glennon, J. C., & Staal, W. G. (2020). Self-initiations in young children with autism during pivotal response treatment with and without robot assistance. *Autism*, 24(8), 2117–2128. <https://doi.org/10.1177/1362361320935006>
- DiStefano, C., Shih, W., Kaiser, A., Landa, R., & Kasari, C. (2016). Communication growth in minimally verbal children with ASD: The importance of interaction. *Autism Research*, 9(10), 1093–1102. <https://doi.org/10.1002/aur.1594>
- Dykstra, J. R., Sabatos-DeVito, M. G., Irvin, D. W., Boyd, B. A., Hume, K. A., & Odom, S. L. (2013). Using the language environment analysis (LENA) system in preschool classrooms with children with autism spectrum disorders. *Autism*, 17(5), 582–594. <https://doi.org/10.1177/1362361312446206>
- Elder, J. H., Donaldson, S. O., Kairalla, J., Valcante, G., Bendixen, R., Ferdig, R., Self, E., Walker, J., Palau, C., & Serrano, M. (2011). In-home training for fathers of children with autism: A follow up study and evaluation of four individual training components. *Journal of Child and Family Studies*, 20(3), 263–271. <https://doi.org/10.1007/s10826-010-9387-2>
- Farmer, C. A., Kaat, A. J., Thurm, A., Anselm, I., Akshoomoff, N., Bennett, A., Berry, L., Bruchey, A., Barshop, B. A., Berry-Kravis, E., Bianconi, S., Cecil, K. M., Davis, R. J., Ficioglu, C., Porter, F. D., Wainer, A., Goin-Kochel, R. P., Leonczyk, C., & Guthrie, W., ... & J. S. Miller (2020). Person ability scores as an alternative to norm-referenced scores as outcome measures in studies of neurodevelopmental disorders. *American Journal on Intellectual and Developmental Disabilities*, 125(6), 475–480. <https://doi.org/10.1352/1944-7558-125.6.475>
- Fasano, R. M., Mitsven, S. G., Custode, S. A., Sarker, D., Bulotsky-Shearer, R. J., Messinger, D. S., & Perry, L. K. (2023). Automated measures of vocal interactions and engagement in inclusive preschool classrooms. *Autism Research*, 16(8), 1586–1599. <https://doi.org/10.1002/aur.2980>
- Fasano, R. M., Perry, L. K., Zhang, Y., Vitale, L., Wang, J., Song, C., & Messinger, D. S. (2021). A granular perspective on inclusion: Objectively measured interactions of preschoolers with and without autism. *Autism Research*, 14(8), 1658–1669. <https://doi.org/10.1002/aur.2526>
- Feldman, H. M., Campbell, T. F., Kurs-Lasky, M., Rockette, H. E., Dale, P. S., Colborn, D. K., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development*, 76(4), 856–868. <https://doi.org/10.1111/j.1467-8624.2005.00882.x>
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), Article i–185. <https://doi.org/10.2307/1166093>
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates communicative*

- development inventories: User's guide and technical manual.* Brookes.
- Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, 10(3), 384–407. <https://doi.org/10.1002/aur.1678>
- Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition*, 183, 1–18. <https://doi.org/10.1016/j.cognition.2018.10.022>
- Fusaroli, R., Weed, E., Rocca, R., Fein, D., & Naigles, L. (2023). Caregiver linguistic alignment to autistic and typically developing children: A natural language processing approach illuminates the interactive components of language development. *Cognition*, 236, Article 105422. <https://doi.org/10.1016/j.cognition.2023.105422>
- Germone, M. M., Gabriels, R. L., Guérin, N. A., Pan, Z., Banks, T., & O'Haire, M. E. (2019). Animal-assisted activity improves social behaviors in psychiatrically hospitalized youth with autism. *Autism*, 23(7), 1740–1751. <https://doi.org/10.1177/1362361319827411>
- Goodwin, A., Fein, D., & Naigles, L. R. (2012). Comprehension of wh-questions precedes their production in typical development and autism spectrum disorders. *Autism Research*, 5(2), 109–123. <https://doi.org/10.1002/aur.1220>
- Gordon, K., Pasco, G., McElduff, F., Wade, A., Howlin, P., & Charman, T. (2011). A communication-based intervention for nonverbal children with autism: What changes? Who benefits? *Journal of Consulting and Clinical Psychology*, 79(4), 447–457. <https://doi.org/10.1037/a0024379>
- Harbison, A. L., Woynaroski, T. G., Tapp, J., Wade, J. W., Warlaumont, A. S., & Yoder, P. J. (2018). A new measure of child vocal reciprocity in children with autism spectrum disorder. *Autism Research*, 11(6), 903–915. <https://doi.org/10.1002/aur.1942>
- Herlihy, L., Knoch, K., Vibert, B., & Fein, D. (2015). Parents' first concerns about toddlers with autism spectrum disorder: Effect of sibling status. *Autism*, 19(1), 20–28. <https://doi.org/10.1177/136236131509731>
- Hilvert, E., Sterling, A., Haebig, E., & Friedman, L. (2020). Expressive language abilities of boys with idiopathic autism spectrum disorder and boys with fragile X syndrome + autism spectrum disorder: Cross-context comparisons. *Autism & Developmental Language Impairments*, 5, Article 239694152091211. <https://doi.org/10.1177/2396941520912118>
- Hudry, K., Smith, J., Pillar, S., Varcin, K. J., Bent, C. A., Boutrus, M., Chetcuti, L., Clark, A., Dissanayake, C., Iacono, T., Kennedy, L., Lant, A., Robinson Lake, J., Segal, L., Slonims, V., Taylor, C., Wan, M. W., Green, J., & Whitehouse, A. J. O. (2023). The utility of natural language samples for assessing communication and language in infants referred with early signs of autism. *Research on Child and Adolescent Psychopathology*, 51(4), 529–539. <https://doi.org/10.1007/s10802-022-01010-3>
- Hughes, M. M., Shaw, K. A., DiRienzo, M., Durkin, M. S., Esler, A., Hall-Lande, J., Wiggins, L., Zahorodny, W., Singer, A., & Maenner, M. J. (2023). The prevalence and characteristics of children with profound autism, 15 sites, United States, 2000–2016. *Public Health Reports*, 138(6), 971–980. <https://doi.org/10.1177/00333549231163551>
- Jones, E. J. H., Gliga, T., Bedford, R., Charman, T., & Johnson, M. H. (2014). Developmental pathways to autism: A review of prospective studies of infants at risk. *Neuroscience & Biobehavioral Reviews*, 39, 1–33. <https://doi.org/10.1016/j.neubiorev.2013.12.001>
- Kaale, A., Smith, L., Nordahl-Hansen, A., Fagerland, M. W., & Kasari, C. (2018). Early interaction in autism spectrum disorder: Mothers' and children's behaviours during joint engagement. *Child: Care, Health and Development*, 44(2), 312–318. <https://doi.org/10.1111/cch.12532>
- Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder: Assessing minimally verbal ASD. *Autism Research*, 6(6), 479–493. <https://doi.org/10.1002/aur.1334>
- Ko, J. A., Miller, A. R., & Vernon, T. W. (2019). Social conversation skill improvements associated with the social tools and rules for teens program for adolescents with autism spectrum disorder: Results of a randomized controlled trial. *Autism*, 23(5), 1224–1235. <https://doi.org/10.1177/1362361318808781>
- Kover, S. T., Choi, E. A., & Naigles, L. R. (2018). *Assessing vocabulary in young children with ASD: Underestimations in parent report compared to spontaneous language samples* [Poster presentation]. The symposium for research in child language disorders. Madison, WI.
- Kover, S. T., Davidson, M. M., Sindberg, H. A., & Weismer, S. E. (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. https://doi.org/10.1044/2014_JSLHR-L-13-0330
- Kozlowski, A. M., Matson, J. L., Horovitz, M., Worley, J. A., & Neal, D. (2011). Parents' first concerns of their child's development in toddlers with autism spectrum disorders. *Developmental Neuropsychology*, 14(2), 72–78. <https://doi.org/10.3109/17518423.2010.539193>
- Laghi, F., Ferri, R., Caramuscio, E., Lonigro, A., Rea, M., & Pezzuti, L. (2022). The floor effect on the ABAS-II in adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 52(11), 4975–4985. <https://doi.org/10.1007/s10803-021-05369-y>
- Lai, P. T. (2020). Expressivity in children with autism and Williams syndrome. *Advances in Autism*, 6(4), 277–288. <https://doi.org/10.1108/AIA-11-2019-0044>
- LeGrand, K. J., Weil, L. W., Lord, C., & Luyster, R. J. (2021). Identifying childhood expressive language features that best predict adult language and communication outcome in individuals with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 64(6), 1977–1991. https://doi.org/10.1044/2021_JSLHR-20-00544
- Lehet, M., Arjmandi, M. K., Houston, D., & Dilley, L. (2021). Circumspection in using automated measures: Talker gender

- and addressee affect error rates for adult speech detection in the Language ENvironment Analysis (LENA) system. *Behavior Research Methods*, 53(1), 113–138. <https://doi.org/10.3758/s13428-020-01419-y>
- Li, X., Peng, Y., & Zheng, X. (2022). The association between mother's descriptive language and children with autism's conversational repair: A moderated mediation analysis. *Journal of Autism and Developmental Disorders*, 54(1), 224–238. <https://doi.org/10.1007/s10803-022-05785-8>
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule*, 2nd ed. Western Psychological Services.
- Luyster, R., Qiu, S., Lopez, K., & Lord, C. (2007). Predicting outcomes of children referred for autism using the MacArthur—Bates communicative development inventory. *Journal of Speech, Language, and Hearing Research*, 50(3), 667–681. [https://doi.org/10.1044/1092-4388\(2007\)047](https://doi.org/10.1044/1092-4388(2007)047)
- MacDonald-Prigent, A., Saiyed, F., Hyde, K., Sharda, M., & Nadig, A. (2023). Response to music-mediated intervention in autistic children with limited spoken language ability. *Journal of Autism and Developmental Disorders*, 54(4), 1438–1452. <https://doi.org/10.1007/s10803-022-05872-w>
- MacFarlane, H., Salem, A. C., Chen, L., Asgari, M., & Fombonne, E. (2022). Combining voice and language features improves automated autism detection. *Autism Research*, 15(7), 1288–1300. <https://doi.org/10.1002/aur.2733>
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Transcription format and programs*, Vol. 1, 3rd ed. (pp. xi, 366). Lawrence Erlbaum Associates Publishers.
- Maenner, M. J. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2020. *MMWR. Surveillance Summaries*, 72, <https://doi.org/10.15585/mmwr.ss7202a1>
- Maes, P., Weyland, M., & Kissine, M. (2023). Describing (pre)linguistic oral productions in 3- to 5-year-old autistic children: A cluster analysis. *Autism*, 27(4), 967–982. <https://doi.org/10.1177/13623613221122663>
- McDaniel, J., & Brady, N. C. (2022). The influence of communication sample length on reliability and convergent validity of vocal measures derived from the communication complexity scale. *Journal of Speech, Language, and Hearing Research*, 65(10), 3881–3889. https://doi.org/10.1044/2022_JSLHR-22-00010
- McDaniel, J., D'Ambrose Slaboch, K., & Yoder, P. (2018). A meta-analysis of the association between vocalizations and expressive language in children with autism spectrum disorder. *Research in Developmental Disabilities*, 72(1), 202–213. <https://doi.org/10.1016/j.ridd.2017.11.010>
- McDaniel, J., Woynarowski, T., Keceli-Kaysili, B., Watson, L. R., & Yoder, P. (2019). Vocal communication with canonical syllables predicts later expressive language skills in preschool-aged children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 62(10), 3826–3833. https://doi.org/10.1044/2019_JSLHR-L-19-0162
- McDaniel, J., Yoder, P., Estes, A., & Rogers, S. J. (2020). Validity of vocal communication and vocal complexity in young children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 50(1), 224–237. <https://doi.org/10.1007/s10803-019-04248-x>
- McDaniel, J., Yoder, P., Estes, A., & Rogers, S. J. (2021). Stability of vocal variables measured during the early communication indicator for children with autism spectrum disorder. *American Journal on Intellectual and Developmental Disabilities*, 126(2), 142–157. <https://doi.org/10.1352/1944-7558-126.2.142>
- McKernan, E. P., Kumar, M., Di Martino, A., Shulman, L., Kolevzon, A., Lord, C., Narayanan, S., & Kim, S. H. (2022). Intra-topic latency as an automated behavioral marker of treatment response in autism spectrum disorder. *Scientific Reports*, 12(1), Article 3255. <https://doi.org/10.1038/s41598-022-07299-w>
- McMahon, C. M., Vismara, L. A., & Solomon, M. (2013). Measuring changes in social behavior during a social skills intervention for higher-functioning children and adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 43(8), 1843–1856. <https://doi.org/10.1007/s10803-012-1733-3>
- Miller, J., & Iglesias, A. (2012). *Systematic analysis of language transcripts (SALT), Research Version 2012* [Computer software]. SALT Software, LLC.
- Miller, L. E., Perkins, K. A., Dai, Y. G., & Fein, D. A. (2017). Comparison of parent report and direct assessment of child skills in toddlers. *Research in Autism Spectrum Disorders*, 41–42, 57–65. <https://doi.org/10.1016/j.rasd.2017.08.002>
- Milton, D. E. M. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & Society*, 27(6), 883–887. <https://doi.org/10.1080/09687599.2012.710008>
- Morrison, K. E., DeBrabander, K. M., Jones, D. R., Faso, D. J., Ackerman, R. A., & Sasson, N. J. (2020). Outcomes of real-world social interaction for autistic adults paired with autistic compared to typically developing partners. *Autism*, 24(5), 1067–1080. <https://doi.org/10.1177/1362361319892701>
- Nadig, A., Lee, I., Singh, L., Bosshart, K., & Ozonoff, S. (2010). How does the topic of conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. *Neuropsychologia*, 48(9), 2730–2739. <https://doi.org/10.1016/j.neuropsychologia.2010.05.020>
- Naigles, L. R., Cheng, M., Xu Rattanasone, N., Tek, S., Khetrapal, N., Fein, D., & Demuth, K. (2016). "You're telling me!" The prevalence and predictors of pronoun reversals in children with autism spectrum disorders and typical development. *Research in Autism Spectrum Disorders*, 27, 11–20. <https://doi.org/10.1016/j.rasd.2016.03.008>
- Naples, A., Tenenbaum, E. J., Jones, R. N., Righi, G., Sheinkopf, S. J., & Eigsti, I.-M. (2022). Exploring communicative competence in autistic children who are minimally verbal: The low verbal investigatory survey for autism (LVIS). *Autism*, 27(5), 1391–1406. <https://doi.org/10.1177/13623613221136657>

- O'Haire, M. E., McKenzie, S. J., Beck, A. M., & Slaughter, V. (2013). Social behaviors increase in children with autism in the presence of animals compared to toys. *PLoS ONE*, 8(2), Article e57010. <https://doi.org/10.1371/journal.pone.0057010>
- Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., Sigman, M., Steinfeld, M. B., & Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(3), 256–266. <https://doi.org/10.1016/j.jaac.2009.11.009>
- Parish-Morris, J., Liberman, M. Y., Cieri, C., Herrington, J. D., Yerys, B. E., Bateman, L., Donaher, J., Ferguson, E., Pandey, J., & Schultz, R. T. (2017). Linguistic camouflage in girls with autism spectrum disorder. *Molecular Autism*, 8(1), Article 48. <https://doi.org/10.1186/s13229-017-0164-6>
- Parladé, M. V., & Iverson, J. M. (2015). The development of coordinated communication in infants at heightened risk for autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 45(7), 2218–2234. <https://doi.org/10.1007/s10803-015-2391-z>
- Pasco, G., Gordon, R. K., Howlin, P., & Charman, T. (2008). The classroom observation schedule to measure intentional communication (COSMIC): An observational measure of the intentional communication of children with autism in an unstructured classroom setting. *Journal of Autism and Developmental Disorders*, 38(10), 1807–1818. <https://doi.org/10.1007/s10803-008-0569-3>
- Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry*, 52(5), 588–598. <https://doi.org/10.1111/j.1469-7610.2010.02332.x>
- Perry, A., Flanagan, H. E., Dunn Geier, J., & Freeman, N. L. (2009). Brief report: The vineland adaptive behavior scales in young children with autism spectrum disorders at different cognitive levels. *Journal of Autism and Developmental Disorders*, 39(7), 1066–1078. <https://doi.org/10.1007/s10803-009-0704-9>
- Plate, S., Yankowitz, L., Resorla, L., Swanson, M. R., Meera, S. S., Estes, A. M., Marrus, N., Cola, M., Petrulla, V., Faggen, A., Pandey, J., Paterson, S., Pruett, J. R., Hazlett, H. C., Dager, S. R., St. John, T., Botteron, K., Zwaigenbaum, L., & Piven, J., ... & J. Parish-Morris (2021). Infant vocalizing and phenotypic outcomes in autism: Evidence from the first 2 years. *Child Development*, 93(2), 468–483. <https://doi.org/10.1111/cdev.13697>
- Rifai, O. M., Fletcher-Watson, S., Jiménez-Sánchez, L., & Crompton, C. J. (2022). Investigating markers of rapport in autistic and nonautistic interactions. *Autism in Adulthood*, 4(1), 3–11. <https://doi.org/10.1089/aut.2021.0017>
- Rollins, P. R., De Froy, A. M., Gajardo, S. A., & Brantley, S. (2022). Pragmatic contributions to early vocabulary and social communication in young autistic children with language and cognitive delays. *Journal of Communication Disorders*, 99, Article 106243. <https://doi.org/10.1016/j.jcomdis.2022.106243>
- Song, A., Cola, M., Plate, S., Petrulla, V., Yankowitz, L., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2021). Natural language markers of social phenotype in girls with autism. *Journal of Child Psychology and Psychiatry*, 62(8), 949–960. <https://doi.org/10.1111/jcpp.13348>
- Song, X.-K., & So, W.-C. (2022). The influence of child-based factors and parental inputs on expressive language abilities in children with autism spectrum disorder. *Autism*, 26(6), 1477–1490. <https://doi.org/10.1177/13623613211054597>
- Stone, W. L., Coonrod, E. E., Turner, L. M., & Pozdol, S. L. (2004). Psychometric properties of the STAT for early autism screening. *Journal of Autism and Developmental Disorders*, 34(6), 691–701. <https://doi.org/10.1007/s10803-004-5289-8>
- 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. <https://doi.org/10.1002/aur.1329>
- Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (1st ed., pp. 335–364). Wiley. <https://doi.org/10.1002/9780470939345.ch12>
- Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., Rice, M., Stoel-Gammon, C., Wetherby, A., & 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. [https://doi.org/10.1044/1092-4388\(2009/08-0136\)](https://doi.org/10.1044/1092-4388(2009/08-0136))
- Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, 20(6), Article e12456. <https://doi.org/10.1111/desc.12456>
- Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the “noun bias” in context: A comparison of English and mandarin. *Child Development*, 70(3), 620–635. <https://doi.org/10.1111/1467-8624.00045>
- Thomas, R. P., Wittke, K., Blume, J., Mastergeorge, A. M., & Naigles, L. (2023). Predicting language in children with ASD using spontaneous language samples and standardized measures. *Journal of Autism and Developmental Disorders*, 53(10), 3916–3931. <https://doi.org/10.1007/s10803-022-05691-z>
- Tovar, A. T., Fein, D., & Naigles, L. R. (2015). Grammatical aspect is a strength in the language comprehension of young children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 58(2), 301–310. https://doi.org/10.1044/2014_JSLHR-L-13-0257
- Trembath, D., Westerveld, M. F., Teppala, S., Thirumanickam, A., Sulek, R., Rose, V., Tucker, M., Paynter, J., Hetzroni, O., Keen, D., & Vivanti, G. (2019). Profiles of vocalization change in children with autism receiving early intervention.

- Autism Research*, 12(5), 830–842. <https://doi.org/10.1002/aur.2075>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., & Garritty, C., ..., & S. E. Straus (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Unwin, K. L., Powell, G., & Jones, C. R. (2022). The use of multi-sensory environments with autistic children: Exploring the effect of having control of sensory changes. *Autism*, 26(6), 1379–1394. <https://doi.org/10.1177/13623613211050176>
- Usher, L. V., Burrows, C. A., Schwartz, C. B., & Henderson, H. A. (2015). Social competence with an unfamiliar peer in children and adolescents with high functioning autism: Measurement and individual differences. *Research in Autism Spectrum Disorders*, 17, 25–39. <https://doi.org/10.1016/j.rasd.2015.05.005>
- Venker, C. E., Bolt, D. M., Meyer, A., Sindberg, H., Ellis Weismer, S., & Tager-Flusberg, H. (2015). Parent telegraphic speech use and spoken language in preschoolers with ASD. *Journal of Speech, Language, and Hearing Research*, 58(6), 1733–1746. https://doi.org/10.1044/2015_JSLHR-L-14-0291
- Veytsman, E., Baker, E., Martin, A. M., Choy, T., Blacher, J., & Stavropoulos, K. (2023). Perceived and observed treatment gains following PEERS: A preliminary study with Latinx adolescents with ASD. *Journal of Autism and Developmental Disorders*, 53(3), 1175–1188. <https://doi.org/10.1007/s10803-022-05463-9>
- Wang, Y., Hartman, M., Aziz, N. A. A., Arora, S., Shi, L., & Tunison, E. (2017). A systematic review of the use of LENA technology. *American Annals of the Deaf*, 162(3), 295–311. <https://doi.org/10.1353/aad.2017.0028>
- Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, 25(7), 1314–1324. <https://doi.org/10.1177/0956797614531023>
- Wetherby, A. M., Watt, N., Morgan, L., & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. *Journal of Autism and Developmental Disorders*, 37(5), 960–975. <https://doi.org/10.1007/s10803-006-0237-4>
- Winder, B. M., Wozniak, R. H., Parladé, M. V., & Iverson, J. M. (2013). Spontaneous initiation of communication in infants at low and heightened risk for autism spectrum disorders. *Developmental Psychology*, 49(10), 1931–1942. <https://doi.org/10.1037/a0031061>
- Woynaroski, T., Watson, L., Gardner, E., Newsom, C. R., Keceli-Kaysili, B., & Yoder, P. J. (2016). Early predictors of growth in diversity of key consonants used in communication in initially preverbal children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 46(3), 1013–1024. <https://doi.org/10.1007/s10803-015-2647-7>
- Xu, Y. A., Naigles, L. R., & Su, Y. E. (2022). Early word order usage in preschool mandarin-speaking typical children and children with autism spectrum disorder: Influences of caregiver input? *Frontiers in Psychology*, 12, Article 766133. <https://doi.org/10.3389/fpsyg.2021.766133>
- Xu, D., Yapanel, U., Gray, S., & Baer, C. (2009). *The LENA TM language environment analysis system: The Interpreted Time Segments (ITS) file*. https://www.lena.org/wp-content/uploads/2016/07/LTR-04-2_ITS_File.pdf
- Yankowitz, L., Schultz, R. T., & Parish-Morris, J. (2019). Pre- and paralinguistic vocal production in ASD: Birth through school age. *Current Psychiatry Reports*, 21(12), 1–22. <https://doi.org/10.1007/s11920-019-1113-1>
- Yoder, P. J., Oller, D. K., Richards, J. A., Gray, S., & Gilkerson, J. (2013). Stability and validity of an automated measure of vocal development from day-long samples in children with and without autism spectrum disorder. *Autism Research*, 6(2), 103–107. <https://doi.org/10.1002/aur.1271>