



Published in final edited form as:

*Augment Altern Commun.* 2023 March ; 39(1): 33–44. doi:10.1080/07434618.2022.2120071.

## Review of methods for conducting speech research with minimally verbal individuals with autism spectrum disorder

Karen V. Chenausky<sup>a,b,c</sup>, Marc Maffei<sup>a</sup>, Helen Tager-Flusberg<sup>c</sup>, Jordan R. Green<sup>a,d</sup>

<sup>a</sup>MGH Institute of Health Professions, Charlestown, MA, USA

<sup>b</sup>Harvard Medical School, Boston, MA, USA

<sup>c</sup>Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA

<sup>d</sup>Speech and Hearing Biosciences and Technology Program, Harvard University, Boston, MA, USA

### Abstract

The purpose of this paper was to review best-practice methods of collecting and analyzing speech production data from minimally verbal autistic speakers. Data on speech production data in minimally verbal individuals are valuable for a variety of purposes, including phenotyping, clinical assessment, and treatment monitoring. Both perceptual (“by ear”) and acoustic analyses of speech can reveal subtle improvements as a result of therapy that may not be apparent when correct/incorrect judgments are used. Key considerations for collecting and analyzing speech production data from this population are reviewed. The definition of “minimally verbal” that is chosen will vary depending on the specific hypotheses investigated, as will the stimuli to be collected and the task(s) used to elicit them. Perceptual judgments are ecologically valid but subject to known sources of bias; therefore, training and reliability procedures for perceptual analyses are addressed, including guidelines on how to select vocalizations for inclusion or exclusion. Factors to consider when recording and acoustically analyzing speech are also briefly discussed. In summary, the tasks, stimuli, training methods, analysis type(s), and level of detail that yield the most reliable data to answer the question should be selected. It is possible to obtain rich high-quality data even from speakers with very little speech output. This information is useful not only for research but also for clinical decision-making and progress monitoring.

### Keywords

Augmentative and alternative communication; autism spectrum disorder; childhood apraxia of speech; minimally verbal; motor speech disorders; speech production

---

The past decade has seen a surge of interest in research on spoken communication in children with neurodevelopmental disorders. This increase may not be surprising: speech

---

**CONTACT** Karen V. Chenausky kvchenausky@mghihp.edu MGH Institute of Health Professions, 36, 1st Ave., Charlestown, MA 02129, USA.

Disclosure statement

No potential conflict of interest was reported by the author(s).

disorders affect a high proportion of children with neurodevelopmental conditions. Indeed, of the 302 genetic syndromes described in Shprintzen (1997), speech is disordered at least some of the time in 235 syndromes (77.8%). One major challenge in conducting research in this population, however, is that many children with neurodevelopmental disorders produce very little spoken language. For example, up to approximately 33% of children with autism spectrum disorder are minimally verbal. Many also experience significant behavioral challenges that make speech and data collection difficult (Tager-Flusberg & Kasari, 2013). Researchers have therefore begun to develop experimental paradigms for conducting research with minimally verbal children with ASD, many of whom are also intellectually disabled, that maximize both their comfort in the lab or clinic and the likelihood of collecting high-quality data (Tager-Flusberg et al., 2017). This paper adds to the discussion by reviewing a series of methods to use and decisions to make, based both on experience and the extant literature, when collecting data to answer pressing questions about speech performance in minimally verbal children with autism. These data can be collected for research as well as for clinical applications such as assessment of the appropriateness of augmentative and alternative communication (AAC) use or treatment progress monitoring. The overarching goal of this article is to lay out best practices and guidelines for sampling and assessing speech production data in minimally verbal children with autism. To that end, it begins by reviewing how different definitions of minimally verbal may affect the study population, then briefly discusses pediatric speech sound disorders (a common comorbidity in minimally verbal children). Next comes a discussion of different stimuli and how to select vocalizations for analysis. Finally, the paper addresses elicitation procedures and discusses how to conduct both perceptual and acoustic analyses.

## The importance of studying speech production in minimally verbal populations

Why is it important to study speech production in children whose speech development is severely limited? First, and foremost, these children provide a unique and valuable perspective on speech development and speech production. Both their challenges and their preserved abilities provide important information that can illuminate the neurology and genetics of speech production. In addition, a deep understanding of verbal communication profiles is important for improving communication therapy, even if the best therapeutic choice is a non-speech augmentative and alternative communication (AAC) modality such as manual sign or a picture- or text-based modality such as a speech-generating device.

Second, speech development is interlinked with language development; thus, understanding how speech and language development are intertwined in children with concomitant disorders of speech and language can lead to important insights into the neural and behavioral bases of these challenges. For example, there are high rates of comorbidity between developmental language disorders and speech disorders such as childhood apraxia of speech (Chenausky et al., 2019, 2021; Tierney et al., 2015; Velleman & Strand, 1994). Chenausky et al. (2019) found that speech-production ability accounted for approximately 33% of the variance in the number of different words produced during a spontaneous language sample for low- and minimally-verbal children with autism and suspected

childhood apraxia of speech and those with insufficient speech output to rate. This line of research also offers hope that speech characteristics such as early consonant inventory may act as predictors of later language outcomes (Yoder et al., 2015). Such insights can help clinicians focus their energy in the most appropriate direction, such as early interventions for increasing babbling frequency and diversity in children at risk for being minimally verbal (Peter et al., 2019) and whether those interventions should include an augmentative or alternative communication component (Millar et al., 2006; Schlosser & Wendt, 2008a, 2008b).

A third reason that a researcher may wish to assess speech in minimally verbal children is to measure baseline skills and monitor progress during interventions designed to improve spoken language production. Such studies include AAC intervention studies that use speech-generating devices to aid the acquisition of spoken language (e.g., Alzrayer et al., 2021; Kasari et al., 2014), treatments that directly address speech (e.g., Chenausky et al., 2016), or multimodal interventions combining both AAC and speech treatment (e.g., Bishop et al., 2020; Brady et al., 2015; Gevarter & Horan, 2019). In particular, measuring proximal improvements in speech production ability can be especially useful when higher-level outcome measures such as correct word production do not show changes in the short term, in order to decide whether to maintain or change the course of intervention (Brady et al., 2021).

## Definition of minimally verbal

In general, the term *minimally verbal* refers to children whose communicative abilities are severely limited as a result of speech impairment, language disorder, intellectual disability, social cognitive ability, or a combination thereof (Chakrabarti, 2017); thus, the term applies to individuals with a wide variety of skills. While language impairment is, by definition, present in this group, scores on standardized tests of language ability and language variables derived from natural language samples may span a wide range. For example, Chenausky et al. (2019) examined a group of 54 minimally- and low-verbal children with autism, ranging from participants who produced no speech at all to participants who produced some multi-word phrases. Receptive vocabulary raw scores in the group ranged from 0 to 123 words recognized; number of different words produced during a language sample ranged from 0 to 229. In terms of speech production, scores on the first two sections of the Kaufman Speech Praxis Test (KSPT; Kaufman, 1995) ranged from 0 to 98% correct. The liberal inclusion criterion in terms of verbal ability ensured that the sample would show adequate variability to explore potential predictors of expressive vocabulary, which would not have been possible with a more restricted range of number of different words. This example illustrates the importance of selecting a definition of minimally verbal that serves the purposes of the study.

Other definitions of minimally verbal include or are based on the number of functional words that person uses by some measure (e.g., Kasari et al., 2008; Koegel et al., 2009, 2020; Yoder & Stone, 2006). For example, Plesa-Skwerer et al. (2016) based their definition on parent report: Children whose parents reported no consistent use of phrase speech and/or production of fewer than 30 words or phrases were considered minimally verbal. Other studies, such as Norrelgen et al. (2015), employed standardized instruments such as the

Vineland Adaptive Behavior Scales (Sparrow et al., 2005); children who were reported to use between three words and some two-word phrases and whose expressive language age equivalent score was lower than 24 months were considered minimally verbal. These more restrictive definitions of minimally verbal may be more appropriate for clinical purposes.

Note that while intellectual disability (defined as a nonverbal IQ score lower than 70) is common in the minimally-verbal population, it is not universal. Bal et al. (2016) report rates of intellectual disability ranging from 85 to 98% in 257 children with autism over age 6 and with mental age greater than 18 months, defined as minimally verbal by different instruments. Rates of intellectual disability in the minimally verbal population vary depending on which IQ test is used. In the sample reported on in Chenausky et al. (2019), mean nonverbal IQ was 68 but ranged from 30 to 115. Half the sample (27) had nonverbal IQ scores below 70, as measured using the Leiter International Performance Scale-Third Edition (Leiter-3; Roid & Miller, 2013).

Bal et al. (2016) also found that across definitions, approximately half of the children classified as minimally verbal showed verbal and nonverbal skills that were commensurate with each other, while the other half showed higher nonverbal than verbal skills. In general, these authors showed that more stringent definitions of minimally verbal selected for more uniform, but more impaired, subgroups. For all of these reasons, the definition of minimally verbal that is used for a particular project deserves careful consideration because it will affect the composition of the sample. A relatively liberal criterion affords the opportunity and statistical power to explore the factors that account for the variation in verbal ability. A narrower criterion might be selected to reduce unwanted variability, for example, when planning therapy or when studying factors that influence whether children transition from being non- or preverbal to producing some spoken words.

## Pediatric speech disorders

While it is not within the scope of this paper to discuss pediatric speech disorders in detail, a brief summary is in order because of emerging evidence that many children with autism and minimal verbal skills also experience comorbid speech disorders. The American Speech-Language-Hearing Association (ASHA) divides speech disorders into several categories (ASHA, n.d.). Functional speech-sound disorders have no known cause and include articulation errors (such as substitutions of [w] for/r/) and ruled-based phonological errors (such as final consonant deletion). In contrast, organic speech sound disorders are those that result from neurological disorders (e.g., childhood apraxia of speech and childhood dysarthria), structural abnormalities (e.g., cleft palate), or sensory disorders (e.g., hearing impairment). In general, phonological disorders are considered linguistic; articulation disorders, childhood apraxia of speech, and dysarthria are structural or motoric in origin. In childhood apraxia of speech, the ability to plan or sequence the movements for speech is impaired, while dysarthria is a disorder of motor execution. Resources for identifying features of childhood dysarthria are scarce and generally concern speech features associated with specific conditions rather than features independent of etiology. Iuzzini-Seigel et al. (2022), however, have produced a tutorial to aid clinicians and researchers in this challenging task. In addition, van Mourik et al. (1997) and Morgan and Liégeois

(2010) discuss characteristics of different dysarthrias in children, while Haas et al. (2021) document how auditory-perceptual aspects of the speech of children with dysarthria change with development. Lists of common developmental phonological processes may be found in Rvachew and Brosseau-Lapré (2018). Another helpful resource is the list of phonological processes in the Khan-Lewis Phonological Analysis test (Khan & Lewis, 2015).

Because childhood apraxia of speech has been identified as occurring more frequently in children with neurodevelopmental disorders, including autism, than in the general population (Baylis & Shriberg, 2018; Chenausky et al., 2019; Fedorenko et al., 2016; Mei et al., 2018; Morgan et al., 2021; Raca et al., 2013; Shriberg, 2008, Shriberg et al., 2011, Shriberg, Strand, et al., 2019), this disorder is briefly discussed in more detail. The core impairment in childhood apraxia of speech is in planning speech movements in the absence of neuromuscular deficits (American Speech-Language-Hearing Association, 2007). This deficit results in speech that is imprecise, inconsistent, and frequently unintelligible. Childhood apraxia of speech is diagnosed clinically by a speech-language pathologist with specific expertise in pediatric motor speech disorders during a thorough speech-language assessment. The speech-language pathologist identifies discriminative features that include, but are not limited to, the three consensus criteria developed by ASHA: (a) lengthened and disrupted coarticulatory transitions between speech sounds, (b) inappropriate prosody, and (c) inconsistent errors (American Speech-Language-Hearing Association, 2007).

Several researchers have presented more detailed lists of features by which childhood apraxia of speech may be identified (Fedorenko et al., 2016; Iuzzini-Seigel et al., 2015; Shriberg et al., 2017). The list of features by Iuzzini-Seigel et al. includes operationalized definitions of each feature, which is particularly useful for research purposes (Chenausky et al., 2019, 2020, 2021). Strand and McCauley (2019) describe a diagnostic protocol more focused on clinical applications, including testing for stimulability (the degree to which clinician scaffolding can help a child achieve correct speech production).

## Choice of speaking tasks and stimuli

Several sources can be used to gather speech data from minimally-verbal children, including (a) natural language samples, (b) word repetition tasks, and (c) nonword repetition tasks. Additionally, (d) nonspeech oral motor tasks and (e) parent questionnaires can be used to assess the degree of oral praxis problems in this population.

### Natural language samples

Natural language samples, which are recordings of participants' spontaneous language production, possess several advantages over the use of standardized testing or parent report for assessing speech performance. First, they can be administered to any participant, even in heterogeneous samples. This distinguishes them from standardized tests, which are appropriate only for participants within certain age ranges, may be normed on samples that do not include children with autism, and may show floor effects for many minimally verbal children with autism. In addition, natural language samples are not subject to practice effects and so can be administered multiple times as probes during a course of treatment. Measures derived from language samples are more ecologically valid than standardized test scores

and may also be more sensitive to change over the course of a therapeutic trial (Barokova & Tager-Flusberg, 2020). For example, frequency or accuracy of the use of a personalized target word selected for an individual child's therapy program can be ascertained from a natural language sample but not a standardized test. Natural language samples can also reveal improvements in volubility that would not be revealed by standardized tests.

A variety of natural language sample formats can be employed. Parent-child interaction, where a caregiver and child play with a standard set of toys, is commonly used in studies of infants and toddlers. Soft toys that make no noise are preferred for speech samples because they will not obscure as many utterances but they are also generally less interesting to play with. Stoel-Gammon (1987, 1989) proposed that connected speech samples consisting of at least 50 speechlike utterances are not only adequate for speech analyses but also constitute a more valid sample of a child's phonological abilities than single-word naming tasks. Binger et al. (2016) show that reliable expressive language and comprehensibility measures, such as mean length of utterance (in words), mean syllables per utterance, and proportion of comprehensible words, can be extracted from language samples even when children's speech is significantly impaired. The Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2012), the gold standard instrument for autism diagnosis, can also double as a language sample and has the advantage of helping minimize demands on the child and family by serving two purposes. The Adapted ADOS (Bal et al., 2020) is appropriate for minimally verbal individuals over the age of 12. Barokova and Tager-Flusberg (2020) provide guidelines for the use of language sampling in children with autism. They show that data derived from natural language samples can show change over time and that these samples possess other important psychometric properties. Barokova et al. (2020) describe a protocol and fidelity process for obtaining natural language samples from minimally verbal children and adolescents with autism. Broome et al. (2017) propose a set of best-practice clinical and research guidelines for assessing speech in children with autism, including prelinguistic speakers.

### Speech imitation tasks

Despite their convenience and clinical utility, there are also some challenges associated with using natural language samples as speech, rather than language, samples. For example, children may be less likely to produce later-developing phonemes such as /z/, /dʒ/, or /ʃ/ spontaneously than in imitation tasks. Furthermore, many minimally verbal children produce little to no spontaneous speech, or they may produce only unintelligible utterances so that accuracy compared to a target cannot be evaluated; therefore, word- or nonword repetition tasks are another possible data source. These tasks involve prompting the child to repeat a word (e.g., "bye" or "mommy") or a nonword (e.g., "pipo") after the examiner. They have the advantage of assessing a child's ability to at least attempt phonemes, phoneme combinations, or syllable structures that do not appear in their spontaneous speech. A child's stimulability can also be assessed during repetition tasks by providing forms of assistance such as unison production or touch cues on successive attempts.

Imitation even of single syllables thus has potential as a valuable data source for minimally verbal children with autism. For example, a protocol in which children are prompted to

imitate syllables representing the corner vowels of English (/bi/, /bae/, /bu/, /ba/) as well as syllables including a variety of manners of articulation (/bʌ/, /mʌ/, /pʌ/, /wʌ/) can provide information about how well the child is able to produce distinct vowels that require different tongue and lip positions; how well they are able to produce the finely graded lip and jaw movements required to distinguish stops from glides; and, if multiple opportunities are provided, the child's ability to consistently achieve a target over several attempts. Stimuli longer than one syllable will necessarily be more difficult for many minimally verbal children with autism, who may produce even two-syllable stimuli one syllable at a time. Insertion of an inappropriate pause between syllables of a word, referred to as *syllable segregation*, is common in this population (Chenausky et al., 2020) but can only be revealed by presenting stimuli that are bisyllabic or longer. The level of difficulty of stimuli should, therefore, be carefully titrated so as to probe the speech features of interest while not unduly frustrating the participants.

### Nonspeech oral tasks

As with multisyllabic stimuli, most tests of nonspeech oral motor function will involve tasks that may be too challenging for many minimally verbal children with autism. Note that receptive language ability, which varies widely in this population, will also affect participants' ability to comprehend task instructions; thus, commonly used maximum performance tasks such as "repeat the syllable 'pa' as quickly and accurately as you can on one breath" are not feasible for many children with limited language comprehension or speech production ability. Some researchers and clinicians, however, have created adaptations of nonspeech oral motor tasks that are more tractable for young children and for children with limited speech sound repertoires. Rupela et al. (2016) describe the use of such a test, the Language-Neutral Assessment of Motor Speech in Young Children (LAMS), in seven preschool and young school-aged children with Down syndrome. Four of the children were able to repeat the syllable "pa" four times in the context of a play-based task involving a doll representing a father ("papa"). Chenausky et al. (2019) employed the first section of the KSPT, which includes nonspeech oral motor tasks such as opening the mouth or lateralizing the tongue. Of the 11 items in this section of the KSPT, on average participants were able to correctly perform seven of them ( $SD = 3.6$ , range: 0–11). These findings indicate that nonspeech oral motor imitation tasks such as those found in the LAMS or KSPT are feasible for minimally verbal children and can potentially reveal the extent to which these children experience comorbid oral apraxia.

### Deciding what speech behaviors to analyze

A fundamental question that arises when collecting speech data from minimally verbal children is what should count as data? The answer is usually taken for granted when working with typical speakers who produce relatively accurate speech and few or no nonspeech vocalizations but is more complex when dealing with children whose speech is both developing and disordered and who often produce vocalizations that are off-target. There are several factors to consider when deciding what vocalizations to include in a data set, one of which is how speechlike the child's vocal attempts are. Children with or at risk for autism, for example, produce higher rates of nonspeechlike or

nontranscribable vocalizations than typically developing children (Plumb & Wetherby, 2013; Schoen et al., 2011; Sheinkopf et al., 2000). These vocalizations are not always considered communicative, though they may provide information about the physiologic constraints over articulator control, just as aspects of repetitive limb movement such as speed or variability can provide information about later motor development (Kanemaru et al., 2013). While previous research has largely focused on more speechlike vocalizations, the logic being that these are more indicative of a child's ability to develop spoken language, more recent research has begun to address whether less speechlike vocalizations (e.g., moans, squeals) from minimally verbal individuals also possess communicative value (Narain, Johnson, Ferguson et al., 2020a; Narain, Johnson, O'Brien et al., 2020b).

Similarly, the stereotyped or repetitive vocalizations that minimally verbal children with autism frequently produce have been the focus of clinical research focused on improving spoken language in these children (Blanc, 2012; Steigler, 2015). These vocalizations, also referred to as *echolalia* or *scripting*, take the form of immediate or delayed repetition of words or phrases, sometimes clearly articulated and resembling productive language, but sometimes unintelligible and resembling infant babble. Immediate echolalia can be an appropriate response to a speech repetition task, but problems may arise when an examiner suspects that a child's response to a prompt is delayed echolalia; that is, the child is producing a self-stimulatory or scripted vocalization rather than attempting the target. This impression generally arises when the phonetic form of the child's vocalization does not resemble the target or when the child produces the same vocalization repeatedly. Accuracy alone, however, cannot be a criterion on which to judge whether a child's production is an attempt at responding because motor speech disorders such as childhood apraxia of speech are associated with distorted or inconsistent productions of stimuli and perseveration is not uncommon (American Speech-Language-Hearing Association, 2007).

The decision about whether to include repetitive vocalizations as data depends on the study's aims. For example, if the goal is to assess a child's ability to repeat a particular word on request or to use it in a socially appropriate communicative manner, then the researcher may want to exclude repetitive vocalizations. On the other hand, if the goal is to collect information about a child's independent phonemic inventory, repetitive vocalizations may provide useful information about their articulation skills. To minimize examiner bias in selecting vocalizations for analysis, two main options are available. One is simply to accept any vocalization as an attempt at a speech target. If vocalizations are scored as correct or coded in some way, scores for vocalizations (even repetitive ones) that do not match the target will be the same as scores for no-response trials (e.g., incorrect or 0); therefore, the risk of false negatives (missing a vocalization that was in fact a response to the prompt) is essentially nil. Counting a repetitive vocalization as a response may raise the chance of a false positive, but exact matches will be rare and will depend heavily on what aspect of the vocalization is being scored or coded. The other option is to create inclusion and exclusion criteria for the vocalizations that will count as data, code them as such, and employ traditional perceptual analysis techniques such as coder training, consensus or independent judgment, and inter-rater reliability measurement (discussed in more detail in the section that follows) to optimize the quality of the data.

Take the example of an intervention study testing a therapy whose aim is to increase the number of different consonants a child produces correctly during a word imitation task. If some proportion of the child's responses to prompts are not clearly imitations of the words but are repetitive vocalizations like "digadigadiga", one option would be to assume that these vocalizations do not constitute attempts at the target and to exclude them from analysis. In this case, raters should train to reliability on identifying the responses that will be counted and then analyze only those utterances. On the other hand, the researcher may choose to accept any vocalization from the child as an attempt at the target. If the rate of repetitive vocalizations is similar at baseline and post-treatment, the chance of one inadvertently being counted as an intentional response to a prompt is essentially the same at both timepoints; therefore, even if some repetitive vocalizations are counted as responses both pre- and post-treatment, this option will not inflate participants' change scores. Another strategy for minimizing the risk of classifying repetitive vocalizations as responses to prompts is to offer the child multiple opportunities to produce the target and select the best token to be scored, however, that is defined in the context of the study (e.g., the token closest to the target by some measure of phonetic accuracy). Of course, in live situations such as during a treatment session, an immediate decision about whether a child's production is accurate enough to reach a criterion may be needed. In this case, assessors must use their best judgment.

## Procedures for optimizing participant performance

Tager-Flusberg et al. (2017) detail a series of methods that can help minimally verbal children with autism perform to the best of their ability on standardized tests and electroencephalography experiments. Three important principles are to communicate to participants what is to happen as clearly as possible, to allow them to move at their own speed, and to provide positive reinforcement for the desired behavior. A visual schedule is useful for explaining to participants what will happen and when. The activities should be explained simply. For example, using *first/then* language (i.e., saying "first work, then break" while pointing to relevant images) is helpful for many children. Even if a participant's language comprehension is low, employing a calm, positive tone will help them feel comfortable.

During the activity, best performance can be encouraged by providing intermittent positive reinforcement. Encouragement, praise, and a compassionate demeanor are important because speaking is very challenging for these children. If needed, the number of trials or tasks left can be indicated visually, for example, by use of a visual countdown board that helps participants keep track of items that remain to be completed (Schlosser et al., 2013). After a specified number of items (say, five) has been completed, the participant will receive a reward. Rewards can take the form of a break from adult attention, time with a preferred toy or other object (as long as it does not become a focus of obsession and interfere with the participant's performance), or a preferred snack.

Moving at a participant's pace means providing breaks and reinforcement as needed for the participant to complete the tasks expeditiously but not hurriedly, and it also means allowing them adequate time to respond to adult prompts. Children with neurodevelopmental

disorders may have more variable response latencies than typically developing children. Negative latencies (when the child begins producing their response before the adult has finished) may occur in the responses of children with childhood apraxia of speech. These children may be unconsciously attempting unison production with the adult to facilitate their own speech production (Strand et al., 2006). Large positive latencies (when the child begins producing their response only after a long delay) may be common in children with intellectual disability, who may simply require time to formulate motor or cognitive responses (Johnson & Parker, 2013). Good general practice is to allow the child adequate time to respond without overwhelming them with commands and to quickly stop speaking when the child does respond so as not to obscure the data on audio recordings.

## Analysis types

Speech data can be analyzed perceptually (i.e., by ear), acoustically (through the use of computerized tools), or by a combination of methods. Perceptual analyses are by far the more common and are appropriate for gathering phoneme-, word-, or word-approximation-level information. Acoustic analyses are more fine-grained and provide information that is not perceptible to the ear but may reveal subtle changes in production. On the other hand, acoustic measures are often less directly clinically interpretable. When combined, however, acoustic and perceptual data can lead to useful and important findings. For example, Macken and Barton (1980) studied the acquisition of the voiced and voiceless stop consonant distinction (e.g., /b/ vs. /p/) in typically developing children. They demonstrated that children begin at a stage in which their attempts at both targets are perceptually and acoustically indistinguishable, then move to one in which their attempts at /b/ and /p/ sound the same (like [b]) but are differentiable acoustically. Finally, they progress to a stage at which they are able to produce /b/s and /p/s that are both perceptually and acoustically distinct. Identification of the intermediate stage, called a *covert contrast*, may be especially important for minimally verbal children with autism, who often appear to make little progress toward higher-level measures of accuracy over the course of treatment. If even sub-perceptual progress can be identified, clinicians can have more confidence that the course of treatment is an effective one.

## Considerations for performing perceptual analyses

Perceptual analyses of speech, including phonetic transcriptions, are ecologically valid and provide valuable information about a speaker's speech behaviors. A common initial step in perceptual analysis is to transcribe a child's speech so as to create a record of the relevant aspects of what the child produced. This transcription can then be used for subsequent analyses without the need to reanalyze the original audio. It is therefore important that these transcriptions be valid and accurate. Like any tool, they are subject to certain kinds of error (Kent, 1996). The intent here is not to detail all the challenges with perceptual analyses and phonetic transcriptions, but rather to briefly discuss major issues and point readers toward other useful references. Issues to consider include appropriate coder training, what tools are available to help with analyses, and the level of detail and scoring method to be used.

## Training in transcription and coding

For creating high-quality phonetic transcriptions of disordered and/or developing speech, a basic course in phonetics and advanced training in transcription of disordered speech is recommended. Ladefoged and Johnson (2015) is a good introduction to basic phonetics, including transcription using the International Phonetic Alphabet (IPA). Shriberg, Kent et al. (2019) provide instruction and practice in transcribing both disordered and developing speech. There are fewer resources available for identifying signs of motor speech disorder such as childhood apraxia of speech or childhood dysarthria in children's speech, but the Dynamic Evaluation of Motor Speech Skills test (DEMSS; Strand & McCauley, 2019) includes training videos for identifying signs of childhood apraxia of speech. Organizations such as Medbridge ([www.medbridge.com](http://www.medbridge.com)) and [SpeechPathology.com](http://www.speechpathology.com) ([www.speechpathology.com](http://www.speechpathology.com)) offer subscription-based online courses on differential diagnosis of pediatric speech disorders that may be useful.

## Transcription and coding tools

Phonetic transcriptions are generally made using pen and paper then typed into an electronic document; this is often fastest for the experienced transcriber. Müller (2006) describes extensions to the International Phonological Alphabet for use in disordered speech and discusses methods for transcribing other aspects of spoken language production, such as stress/intonation and voice quality. Electronic tools exist as well. An easy-to-install phonetic font, Doulos SIL, is available for free online ([software.sil.org/doulos](http://software.sil.org/doulos)). A "typewriter"-based phonetic transcription system such as Klattese (Vitevitch & Luce, 2004) or ARPABet (<https://nlp.stanford.edu/courses/lsa352/arpabet.html>) may also be used. Tools such as LIPP (Logical International Phonetics Programs; <http://www.ihsys.com/site/LIPP.asp?tab=4>) and Phon ([www.phon.ca](http://www.phon.ca)) permit automated analyses of transcriptions and are available online for pay or free, respectively. Another free tool, PEPPER (Programs to Examine Phonetic and Phonologic Evaluation Records), is available online (<https://phonology.waisman.wisc.edu/about-pepper/>) and is a comprehensive software program that aids clinicians and researchers in selecting and administering speech tasks, performing transcriptions, achieving agreement, and performing a number of analyses on the resulting data. It includes analyses of speech, prosody, and voice variables and the ability to print graphs documenting treatment progress. Further information appears in Shriberg, Kent, et al. (2019).

## Level of detail in transcriptions

The level of detail to include in transcriptions is another issue to consider. In general, the tradeoff is between level of detail and ease of achieving reliability. The narrower the transcription, the more information there is about what the speaker is doing but the harder it will be to achieve intra- and inter-rater reliability (Shriberg & Lof, 1991). The Child Speech Disorder Research Network has published a set of guidelines for producing hand transcriptions (Cleland et al., 2017), including a helpful decision tree for deciding what level of detail to employ. In addition, their good practice guidelines include examples of perceptually-based phonological analyses that are in common use clinically (Bates & Titterton, 2017). Stemberger and Bernhardt (2020) also provide a brief tutorial in transcription of child speech.

### Minimizing bias in transcriptions

More than one rater or transcriber should be employed to minimize bias. Training in identifying the perceptual dimension to be coded is necessary to ensure that raters are consistent with each other; this applies whether the goal is phonetic transcription or coding for other perceptual features such as signs of childhood apraxia of speech. Statistics for evaluating reliability vary depending on the measure used. Percent reliability, which captures the proportion of items rated by different judges that were judged the same, was formerly common but is being replaced by measures which take into account chance agreement, such as Cohen's  $\kappa$  and intra-class coefficients (ICCs) (Chenausky et al., 2020; Iuzzini-Seigel et al., 2017). The overview and tutorial on inter-rater reliability in Hallgren (2012) is a good general introduction to the subject. Koo and Li (2016) provide useful guidelines and a flowchart for selecting the correct ICC model for a particular research question.

### Achieving reliability on transcriptions and codes

Challenges to achieving reliability on transcriptions include how to assign equivalences between different transcriptions and how to tally errors in children's speech. If reliability on phonetic transcriptions per se is the goal, Oller and Ramsdell (2006) recommend the use of a weighted reliability measure that takes into account instances where there is not an absolute match between two transcriptions, such as when one transcriber may indicate lip rounding on the initial nasal consonant in "mommy" using a diacritic ([<sub>lip</sub>]), while another may indicate it by inserting a "w" between the first two phonemes ([mwami]). The method proposed by Oller and Ramsdell weights discrepancies according to established phonological principles, so that transcriptions that phonologically similar to each other (as in the example above) are weighted differently than those that are phonologically discrepant.

There is no commonly accepted criterion level of agreement for transcription of disordered or developing speech, and figures reported in the literature span a wide range. Agreement of 85% or greater on point-to-point transcriptions (Pye et al., 1988) is a worthwhile if ambitious goal. Shriberg and Lof (1991) report a range of 20%–100% agreement on transcription of speech samples from typical and disordered speakers, depending on how the samples were obtained and how reliability was assessed. In their landmark study of infant babbles, Davis and MacNeilage (1995) report a mean percent reliability of approximately 45% (range: 33%–69%) for vowel transcription; for consonants, mean percent reliability was approximately 77% (range: 63%–83%). More recently, Oller and Ramsdell (2006) report a mean percent reliability of 21% (range: 0%–59%) across phonemes, which increases to a mean of 60% (range: 33%–85%) with weighting, on transcriptions of infant babbles.

Because of the challenges of achieving reliability on transcriptions themselves, it is generally advisable to aim for reliability on the specific aspect of the transcription that is germane to the research question. For example, was the token produced correctly or not? Was a certain feature present in the token? Researchers should aim to achieve reliability on the same variable that will be used in their analysis. This approach has the advantages of being (a) less labor intensive, (b) able to focus effort on the specific level of analysis that will matter to the statistics, and (c) easier to achieve high reliability. Rvachew et al. (2002)

demonstrate that such coding of infant vocalizations can result in inter-rater reliability figures of at least 80% and Cohen's  $\kappa$  values of at least 0.733 – higher values than those for phonetic transcriptions. The Weighted Speech Sound Accuracy measure developed by Preston et al. (2011) adapts the weighted reliability measure of Oller and Ramsdell (2006) to scoring children's productions of known targets, when documenting accuracy is the goal.

An alternative to transcription-based analyses is to use visual analog scales for rating children's speech productions. Listeners using these scales are presented with a display consisting of a line whose endpoints represent the extremes of the dimension to be rated (e.g., 0–100% intelligible) and are asked to indicate on the line where they believe the child's production to fall. These ratings can be generated using special software to automatically measure the location the listener has chosen, but a pencil-and-paper version, where listeners make a tick mark on a line of known length, can be just as useful. In this case the distance between one endpoint and the tick mark would constitute the score. Munson et al. (2012) showed that both of these methods are reliable and sensitive enough to indicate change over time in the speech of children receiving speech treatment.

A final issue concerns the method used to achieve reliability or to create the transcription of record (i.e., the “official” transcription of the child's utterance that will then be used in analyses going forward). Transcribers typically receive accuracy training prior to analyzing data and estimating transcription reliability. The most commonly used method is for new transcribers to train to a criterion level of reliability on an already-transcribed set of recordings that is similar to the one being investigated but which will not be used for the research question. Then, two or more transcribers independently transcribe a subset of the recordings (generally at least 10%) that will be used for the research question. Reliability is calculated from these independent transcriptions. Once an acceptable level of reliability has been achieved (after repeating the previous steps if necessary to achieve adequate agreement), one transcriber can then transcribe the remainder of the recordings. This method is relatively efficient but incorporates only one listener's judgments going forward. Consensus methods of transcription and the use of average scores reduce this type of bias by incorporating information from more than one transcriber, but they are more time- and labor intensive because the entire data set must be transcribed by multiple listeners. Finally, decisions about how to handle utterances on which consensus was difficult or impossible can introduce other sources of bias into the study. To address these problems, Shriberg et al. (1984) provide a set of procedures for generating consensus transcriptions, many of which are useful when consensus-coding utterances as well.

### **Considerations for performing acoustic analyses**

It is not within the scope of this article to provide an indepth tutorial on how to obtain high-quality audio recordings or make acoustic measurements. Vogel and Maruff (2008), however, compare commonly available methods of acquiring acoustic data against industry-standard methods and equipment, showing that many measures, especially those involving time-domain measurements such as those used on the /b/ and /p/ tokens in Macken and Barton (1980), can be accurately extracted even with consumer laptop computers. Though Vogel and Maruff used adult speakers who could maintain a constant distance from

the microphone, time-domain measurements like Macken and Barton's can be robustly measured from audio recordings made without constraining mouth-to-microphone distance. In Chenausky and Tager Flusberg (2017), for instance, between-judge measurement reliability of voice onset time (a main acoustic difference between /b/ and /p/) from audio files of ADOS assessments, where children are free to move about the room, was high (Pearson's  $r=0.902$ ,  $p < 0.05$ ) and mean difference in voice onset time between judges was 0.6 ms, compared to 6 ms (Macken & Barton, 1980) and 2.1 ms (Hitchcock & Koenig, 2013) in other studies.

Another aspect of speech that has been measured in typically-developing populations is *vowel formants*, or the resonant frequencies of the vocal tract that make /i /sound different from /a/, for example (Vorperian & Kent, 2007). Formant frequency measurements have been used to document the degree to which children are able to make phonetic distinctions between vowels, and these distinctions have been shown to be reduced and related to reduced intelligibility in children with dysarthria (Allison et al., 2017; Higgins & Hodge, 2002). Several free acoustic analysis programs are available that include automatic formant measurement capabilities: Praat (Boersma & Weenink, 2010), TF32 (Milenkovic, 2010), and Wavesurfer (Sjolander & Beskow, 2019). Derdemezis et al. (2016) compared formant values from each of these programs, measured in the speech of children with Down syndrome, to values from manual measurement obtained via a consensus analysis procedure, and described how to optimize parameter settings in each one to maximize measurement accuracy for children's speech. Kent et al. (2010) also provide valuable suggestions on how to improve or maximize acoustic analyses of children's speech. Note that special considerations apply to audio recordings made over the internet because the file formats employed by many video conferencing programs reduce data quality. See Zhang et al. (2021) for details regarding remote recording and analysis of speech data. Finally, a normalization procedure must be employed when comparing vowel formants across children of different ages because vocal tract size (which increases with age) affects formant values. Fourakis et al. (1993) document one such normalization procedure.

## Discussion

This paper describes a series of methods and considerations for collecting and analyzing speech data from minimally verbal children with autism for both research and clinical purposes. Those purposes include phenotyping in neurodevelopmental disorders, testing speech treatment efficacy, and determining the efficacy of augmentative or alternative communication approaches for increasing spoken language production. Careful perceptual and acoustic analyses of the speech of children with minimal verbal skills can reveal details about potential responses to therapies that simple correct/incorrect judgements of whole words cannot.

Choosing a definition of minimally verbal that will allow for appropriate quantitative variation in performance on the variables of interest, while controlling for comorbidities that might introduce unwanted variation in a cohort, is an important initial step. Selection of task and stimuli is equally important. Natural language samples have high ecological validity but may not result in much data from children with extremely limited vocal output. Speech

and nonspeech repetition tasks can yield much useful information for children who do not produce much spontaneous speech.

The data analysis method should also be selected with care and consideration of both levels of detail and reliability needed and time and cost. In general, finer levels of detail, such as those associated with narrow phonetic transcription, make it harder to achieve adequate levels of reliability. Yet, simple, time-efficient measures such as visual analog scale ratings obtained from children's responses on standardized tests or other tasks also show very high correlation with more detailed measures such as percent phonemes correct and can yield rich data that is adequate for many clinical applications.

Making a principled decision about what kinds of vocalizations to include or exclude to answer a specific clinical or research question is important for obtaining clean data. So is employing assessment techniques that involve clear communication about task requirements that is appropriate to a child's comprehension level, combined with positive reinforcement for desired behaviors and respect for a child's pace and limits.

Both perceptual and acoustic analyses of speech yield rich and detailed information. Perceptual analyses are more ecologically valid and more common clinically but are subject to known sources of bias that should be controlled for by the use of multiple raters. Adequate training in transcription and identification of the desired aspects of speech production will yield the highest reliability. Acoustic analyses of speech production complement perceptual analyses but are likely more useful for research purposes, rather than clinical applications. When performing acoustic analyses, careful attention to recording environment, audio equipment and file formats will yield the highest-quality data.

Perhaps the most important take-away is to identify the level of detail for one's clinical or research purposes and select the tasks, stimuli, training methods, and analyses that yield the most reliable data to answer the questions. Even children who produce very little speech are capable of showing us a wide range of communicative behaviors that are important not only for tracking progress in therapy but also selecting a communication modality (e.g., speech, sign, picture) that allows them to make their needs and wants known and at the same time maximizes their ability to acquire as much receptive and expressive language as possible.

## Funding

This work was supported by National Institute of Deafness and Other Communication Disorders under [NIH P50 DC 018006 (PI HTF, also supporting KVC, JRG, and MM), K24 DC 016312 (PI JRG), and R00 DC 017490 (PI KVC)].

## References

- Allison KM, Annear L, Policicchio M, & Hustad KC (2017). Range and precision of formant movement in pediatric dysarthria. *Journal of Speech, Language, and Hearing Research*, 60(7), 1864–1876. doi:10.1044/2017\_JSLHR-S-15-0438
- Alzrayer NM, Aldabas R, Alhossein A, & Alharthi H (2021). Naturalistic teaching approach to develop spontaneous vocalizations and augmented communication in children with autism spectrum disorder. *Augmentative and Alternative Communication*, 37(1), 14–24. doi:10.1080/07434618.2021.1881825 [PubMed: 33825612]

- American Speech-Language-Hearing Association. (2007). Childhood apraxia of speech [Technical Report]. <https://www.asha.org/policy/ps2007-00277/>
- American Speech-Language-Hearing Association. (n.d.). Speech sound disorders-articulation and phonology. <https://www.asha.org/practice-portal/clinical-topics/articulation-and-phonology> (accessed 2022 Jan 31).
- Bal VH, Katz T, Bishop SL, & Krasileva K (2016). Understanding definitions of minimally verbal across instruments: Evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 57(12), 1424–1433. doi:10.1111/jcpp.12609 [PubMed: 27473432]
- Bal VH, Maye M, Salzman E, Huerta M, Pepa L, Risi S, & Lord C (2020). The adapted ADOS: A new module set for the assessment of minimally verbal adolescents and adults. *Journal of Autism and Developmental Disorders*, 50(3), 719–729. doi:10.1007/s10803-019-04302-8 [PubMed: 31736004]
- Barokova M, & Tager-Flusberg H (2020). Commentary: Measuring language change through natural language samples. *Journal of Autism and Developmental Disorders*, 50(7), 2287–2306. doi:10.1007/s10803-018-3628-4 [PubMed: 29873016]
- Barokova M, La Valle C, Hassan S, Lee C, Xu M, McKechnie R, Johnston E, Krol M, Leañó J, & Tager-Flusberg H (2020). Eliciting language samples for analysis (ELSA): A new protocol for assessing expressive language and communication in autism. *Autism Research*, 14(1):112–126. doi:10.1002/aur.2380 [PubMed: 32909382]
- Bates S, Titterton J (2017). Good practice guidelines for the analysis of child speech. Child Speech Disorder Research Network, North Bristol, UK. <https://www.nbt.nhs.uk/bristol-speech-language-therapy-research-unit/bsltru-research/child-speech-disorder-research-network/resources-relating-clinical-management-child-speech-disorder> (accessed 2021 Jul 30).
- Baylis A, & Shriberg L (2018). Estimates of the prevalence of speech and motor speech disorders in youth with 22q11.2 deletion syndrome. *American Journal of Speech-Language Pathology*, 28(1), 53–82. doi:10.1044/2018\_AJSLP-18-0037
- Binger C, Ragsdale J, & Bustos A (2016). Language sampling for preschoolers with severe speech impairments. *American Journal of Speech-Language Pathology*, 25(4), 493–507. doi:10.1044/2016\_AJSLP-15-0100 [PubMed: 27552110]
- Bishop SK, Moore JW, Dart EH, Radley K, Brewer R, Barker L-K, Quintero L, Litten S, Gilfeather A, Newborne B, & Toche C (2020). Further investigation of increasing vocalizations of children with autism with a speech-generating device. *Journal of Applied Behavior Analysis*, 53(1), 475–483. [PubMed: 30900254]
- Blanc M (2012). Natural language acquisition on the autism spectrum: The journey from echolalia to self-generated language. Communication Development Center.
- Boersma P, Weenink D (2010). Praat [Computer software]. <https://www.fon.hum.uva.nl/praat/>
- Brady N, Kosirog C, Fleming K, & Williams L (2021). Predicting progress in word learning for children with autism and minimal verbal skills. *Journal of Neurodevelopmental Disorders*, 13(1), 36. doi:10.1186/s11689-021-09386-x [PubMed: 34525947]
- Brady N, Storkel H, Bushnell P, Barker R, Saunders K, Daniels D, & Fleming K (2015). Investigating a multimodal intervention for children with limited expressive vocabularies associated with autism. *American Journal of Speech-Language Pathology*, 24(3), 438–459. doi:10.1044/2015\_AJSLP-14-0093 [PubMed: 25910710]
- Broome K, McCabe P, Docking K, & Doble M (2017). A systematic review of speech assessments for children with autism spectrum disorder: Recommendations for best practice. *American Journal of Speech-Language Pathology*, 26(3), 1011–1029. doi:10.1044/2017\_AJSLP-16-0014 [PubMed: 28772287]
- Chakrabarti B (2017). Commentary: Critical considerations for studying low-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 58(4), 436–438. doi:10.1111/jcpp.12720 [PubMed: 28346760]
- Chenausky KV, Brignell A, Morgan A, Gagne D, Norton A, Tager-Flusberg H, Schlaug G, Shield A, & Green J (2020). Factor analysis of signs of childhood apraxia of speech. *Journal of Communication Disorders*, 87, 106033. doi:10.1016/j.jcomdis.2020.106033 [PubMed: 32877838]

- Chenausky KV, Brignell A, Morgan A, Norton A, Tager-Flusberg H, Schlaug G, & Guenther FH (2021). A modeling-guided case study of disordered speech in minimally verbal children with autism spectrum disorder. *American Journal of Speech-Language Pathology*, 30(3S), 1542–1557. doi:10.1044/2021\_AJSLP-20-00121 [PubMed: 33852328]
- Chenausky K, & Tager-Flusberg H (2017). Acquisition of voice onset time in toddlers at high and low risk for autism spectrum disorder: VOT acquisition in toddlers at risk for ASD. *Autism Research*, 10(7), 1269–1279. doi:10.1002/aur.1775 [PubMed: 28339140]
- Chenausky K, Brignell A, Morgan A, & Tager-Flusberg H (2019). Motor speech impairment predicts expressive language in minimally verbal, but not low verbal, individuals with autism spectrum disorder. *Autism & Developmental Language Impairments*, 4, 239694151985633. doi:10.1177/2396941519856333
- Chenausky K, Norton A, Tager-Flusberg H, & Schlaug G (2016). Auditory-motor mapping training: Comparing the effects of a novel speech treatment to a control treatment for minimally verbal children with autism. *PLOS One*, 11(11), e0164930. doi:10.1371/journal.pone.0164930 [PubMed: 27829034]
- Cleland J, Titterton J, Stringer H, Rees R (2017). Good practice guidelines for the transcription of children’s speech in clinical practice and research. Child Speech Disorder Research Network. [https://www.nbt.nhs.uk/sites/default/files/BSLTRU\\_Good%20practice%20guidelines\\_Transcription\\_2Ed\\_2017.pdf](https://www.nbt.nhs.uk/sites/default/files/BSLTRU_Good%20practice%20guidelines_Transcription_2Ed_2017.pdf)
- Davis BL, & MacNeilage PF (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, 38(6), 1199–1211. doi:10.1044/jshr.3806.1199 [PubMed: 8747814]
- Derdemezis E, Vorperian HK, Kent RD, Fourakis M, Reinicke EL, & Bolt DM (2016). Optimizing vowel formant measurements in four acoustic analysis systems for diverse speaker groups. *American Journal of Speech-Language Pathology*, 25(3), 335–354. doi:10.1044/2015\_AJSLP-15-0020 [PubMed: 26501214]
- Fedorenko E, Morgan A, Murray E, Cardinaux A, Mei C, Tager-Flusberg H, Fisher SE, & Kanwisher N (2016). A highly penetrant form of childhood apraxia of speech due to deletion of 16p11.2. *European Journal of Human Genetics*, 24(2), 302–306. doi:10.1038/ejhg.2015.149 [PubMed: 26173965]
- Fourakis M, Geers A, & Tobey E (1993). An acoustic metric for assessing change in vowel production by profoundly hearing-impaired children. *The Journal of the Acoustical Society of America*, 94(5), 2544–2552. doi:10.1121/1.407366 [PubMed: 8270732]
- Gevarter C, & Horan K (2019). A behavioral intervention package to increase vocalizations of individuals with autism during speech-generating device intervention. *Journal of Behavioral Education*, 28(1), 141–167. doi:10.1007/s10864-018-9300-4
- Haas E, Ziegler W, & Schölderle T (2021). Developmental courses in childhood dysarthria: Longitudinal analyses of auditory-perceptual parameters. *Journal of Speech, Language, and Hearing Research*, 64(5), 1421–1435. doi:10.1044/2020\_JSLHR-20-00492
- Hallgren KA (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. doi:10.20982/tqmp.08.1.p023 [PubMed: 22833776]
- Higgins C, & Hodge M (2002). Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology*, 20(4), 271–277.
- Hitchcock ER, & Koenig LL (2013). The effects of data reduction in determining the Schedule of voicing acquisition in young children. *Journal of Speech, Language, and Hearing Research*, 56(2), 441–457. doi:10.1044/1092-4388(2012/11-0175)
- Iuzzini-Seigel J, Allison KM, & Stoeckel R (2022). Identifying childhood apraxia of speech and dysarthria in children with complex speech patterns: A tutorial. [Unpublished manuscript]. Speech Pathology and Audiology, Marquette University.
- Iuzzini-Seigel J, Hogan TP, & Green JR (2017). Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language, and Hearing Research*, 60(5), 1194–1210. doi:10.1044/2016\_JSLHR-S-15-0184

- Iuzzini-Seigel J, Hogan TP, Guarino AJ, & Green JR (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders*, 54, 32–42. doi:10.1016/j.jcomdis.2015.01.002 [PubMed: 25662298]
- Johnson N, & Parker A (2013). Effects of wait time when communicating with children who have sensory and additional disabilities. *Journal of Visual Impairment & Blindness*, 107(5), 363–374. doi:10.1177/0145482X1310700505
- Kanemaru N, Watanabe H, Kihara H, Nakano H, Takaya R, Nakamura T, Nakano J, Taga G, & Konishi Y (2013). Specific characteristics of spontaneous movements in preterm infants at term age are associated with developmental delays at age 3 years. *Developmental Medicine & Child Neurology*, 55, n/a–n/a. doi:10.1111/dmcn.12156
- Kasari C, Kaiser A, Goods K, Nietfeld J, Mathy P, Landa R, Murphy S, & Almirall D (2014). Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 53(6), 635–646. doi:10.1016/j.jaac.2014.01.019 [PubMed: 24839882]
- Kasari C, Paparella T, Freeman S, & Jahromi L (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology*, 76(1), 125–137. doi:10.1037/0022-006X.76.1.125 [PubMed: 18229990]
- Kaufman N (1995). *Kaufman speech praxis test*. Wayne State University Press.
- Kent RD (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, 5(3), 7–23. doi:10.1044/1058-0360.0503.07
- Kent R, Pagan-Neves L, Hustad K, & Fiszbein-Wertzner H (2010). Children's speech sound disorders: An acoustic perspective. In Paul R, & Flipsen P (Eds.), *Speech sound disorders in children: In honor of Laurence D. Shriberg*. Plural Publishing.
- Khan L, & Lewis N (2015). *Khan-Lewis phonological analysis*. Pearson Publishing.
- Koegel L, Bryan K, Su P, Vaidya., & Camarata S (2020). Definitions of nonverbal and minimally verbal in research for autism: A systematic review of the literature. *Journal of Autism and Developmental Disorders*, 50(8), 2957–2972. doi:10.1007/s10803-020-04402-w [PubMed: 32056115]
- Koegel RL, Shirotova L, & Koegel LK (2009). Brief report: Using individualized orienting cues to facilitate first-word acquisition in non-responders with autism. *Journal of Autism and Developmental Disorders*, 39(11), 1587–1592. doi:10.1007/s10803-009-0765-9 [PubMed: 19488847]
- Koo TK, & Li MY (2016). A guideline of selecting and reporting intra-class correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. doi:10.1016/j.jcm.2016.02.012 [PubMed: 27330520]
- Ladefoged P, & Johnson K (2015). *A course in phonetics (7th ed.)*. Cengage Learning.
- Lord C, Rutter M, DiLavore P, Risi S, & Gotham K (2012). *Autism diagnostic observation schedule (Modules 1–4) (2nd ed.)*. Western Psychological Services.
- Macken M, & Barton D (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7(1), 41–74. doi:10.1017/S0305000900007029 [PubMed: 7372738]
- Mei C, Fedorenko E, Amor D, Boys A, Hoeflin C, Carew P, Burgess T, Fisher S, & Morgan A (2018). Deep phenotyping of speech and language skills in individuals with 16p11.2 deletion. *European Journal of Human Genetics: EJHG*, 26(5), 676–686. doi:10.1038/s41431-018-0102-x [PubMed: 29445122]
- Milenkovic P (2010). TF32 [Computer software]. <http://userpages.chorus.net/cspeech/>
- Millar D, Light J, & Schlosser R (2006). The impact of augmentative and alternative communication intervention on the speech production of individuals with developmental disabilities: A research review. *Journal of Speech, Language, and Hearing Research: Jslhr*, 49 (2)9, 248–264. doi:10.1044/1092-4388(2006/021) [PubMed: 16671842]
- Morgan A, & Liégeois F (2010). Re-thinking diagnostic classification of the dysarthrias: A developmental perspective. *Folia Phoniatica et Logopaedica*, 62(3), 120–126. doi:10.1159/000287210 [PubMed: 20424467]

- Morgan A, Braden R, Wong MMK, Colin E, Amor D, Liégeois F, Srivastava S, Vogel A, Bizaoui V, Ranguin K, Fisher SE, & van Bon BW (2021). Speech and language deficits are central to SETBP1 haploinsufficiency disorder. *European Journal of Human Genetics*, 29(8), 1216–1225. doi:10.1038/s41431-021-00894-x [PubMed: 33907317]
- Muüller N (2006). *Multilayered transcription*. Plural Publishing, Inc.
- Munson B, Schellinger SK, & Carlson KU (2012). Measuring speech-sound learning using visual analog scaling. *Perspectives on Language Learning and Education*, 19(1), 19–30. doi:10.1044/lle19.1.19
- Narain J, Johnson KT, Ferguson C, O'Brien A, Talkar T, Zhang Weninger Y, Wofford P, Quatieri T, Picard R, & Maes P (2020a). Personalized modeling of real-world vocalizations from nonverbal individuals [Paper presentation]. *Proceedings of the International Conference on Multimodal Interaction (ICMI)*, Utrecht, Netherlands.
- Narain J, Johnson KT, O'Brien A, Wofford P, Maes P, & Picard RW (2020b). Nonverbal vocalizations as speech: Characterizing natural-environment audio from nonverbal individuals with autism [Paper presentation]. *Proceedings of the Workshop on Laughter and Other Nonverbal Vocalisations*, Bielefeld, Germany.
- Norrelgen F, Fernell R, Eriksson M, Hedvall Å, Persson C, Sjölin M, Gillberg C, & Kjellmer L (2015). Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism: The International Journal of Research and Practice*, 19(8), 934–943. doi:10.1177/1362361314556782 [PubMed: 25488002]
- Oller DK, & Ramsdell HL (2006). A weighted reliability measure for phonetic transcription. *Journal of Speech, Language, and Hearing Research* 49(6), 1391–1411. doi:10.1044/1092-4388(2006/100)
- Peter B, Potter N, Davis J, Donenfeld-Peled I, Finestack L, Stoel-Gammon C, Lien K, Bruce L, Vose C, Eng L, Yokoyama H, Olds D, & VanDam M (2019). Toward a paradigm shift from deficit-based to proactive speech and language treatment: Randomized pilot trial of the Babble Boot Camp in infants with classic galactosemia. *F1000Research*, 8, 271–303. doi:10.12688/f1000research.18062.5 [PubMed: 32566130]
- Plesa-Skwerer D, Jordan S, Brukilacchio B, & Tager-Flusberg H (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism: The International Journal of Research and Practice*, 20(5), 591–604. doi:10.1177/1362361315600146 [PubMed: 26408635]
- Plumb AM, & Wetherby AM (2013). Vocalization development in toddlers with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 56(2), 721–734. doi:10.1044/1092-4388(2012/11-0104)
- Preston JL, Ramsdell HL, Oller DK, Edwards ML, & Tobin SJ (2011). Developing a weighted measure of speech sound accuracy. *Journal of Speech, Language, and Hearing Research*, 54(1), 1–18. doi:10.1044/1092-4388(2010/10-0030)
- Pye C, Wilcox K, & Siren K (1988). Refining transcriptions: The significance of transcriber “errors”. *Journal of Child Language*, 15(1), 17–37. doi:10.1017/S0305000900012034 [PubMed: 3280577]
- Raca G, Baas B, Kirmani S, Laffin J, Jackson C, Strand E, Jakielski K, & Shriberg L (2013). Childhood apraxia of speech (CAS) in two patients with 16p11.2 microdeletion syndrome. *European Journal of Human Genetics: EJHG*, 21(4), 455–459. doi:10.1038/ejhg.2012.165 [PubMed: 22909774]
- Roid G, & Miller L (2013). *Leiter-3: Leiter international performance scale (3rd ed.)*. Western Psychological Services.
- Rupela V, Velleman SL, & Andrianopoulos MV (2016). Motor speech skills in children with Down syndrome: A descriptive study. *International Journal of Speech-Language Pathology*, 18(5), 483–492. doi:10.3109/17549507.2015.1112836 [PubMed: 27063682]
- Rvachew S, & Brosseau-Lapré F (2018). *Developmental phonological disorders: Foundations of clinical practice (2nd ed.)*. Plural Publishing, Inc.
- Rvachew S, Creighton D, Feldman N, & Sauve R (2002). Acoustic-phonetic description of infant speech samples: Coding reliability and related methodological issues. *Acoustics Research Letters Online*, 3(1), 24–28. doi:10.1121/1.1429202

- Schlosser RW, & Wendt O (2008b). Effects of augmentative and alternative communication intervention on speech production in children with autism: A systematic review. *American Journal of Speech-Language Pathology*, 17(3), 212–230. doi:10.1044/1058-0360(2008/021) [PubMed: 18663107]
- Schlosser RW, Laubscher E, Sorce J, Koul R, Flynn S, Hotz L, Abramson J, Fadie H, & Shane H (2013). Implementing directives that involve prepositions with children with autism: A comparison of spoken cues with two types of augmented input. *Augmentative and Alternative Communication*, 29(2), 132–145. doi:10.3109/07434618.2013.784928 [PubMed: 23705815]
- Schlosser RW, & Wendt O (2008a). Augmentative and alternative communication intervention for children with autism. In Luiselli J, Russo D, Christian W, & Wliczinski S (Eds.), *Effective practices for children with autism: Educational and behavioral support interventions that work*. Oxford University Press.
- Schoen E, Paul R, & Chawarska K (2011). Phonology and vocal behavior in toddlers with autism spectrum disorders. *Autism Research*, 4(3), 177–188. doi:10.1002/aur.183 [PubMed: 21308998]
- Sheinkopf SJ, Mundy P, Oller DK, & Steffens M (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders*, 30(4), 345–354. doi:10.1023/a:1005531501155 [PubMed: 11039860]
- Shprintzen RJ (1997). *Genetics, syndromes, and communication disorders*. Singular Publishing Group, Inc.
- Shriberg L, Potter N, & Strand E (2011). Prevalence and phenotype of childhood apraxia of speech in youth with galactosemia. *Journal of Speech, Language, and Hearing Research*, 54(2), 487–519. doi:10.1044/1092-4388(2010/10-0068)
- Shriberg L (2008). Childhood apraxia of speech (CAS) in neurodevelopmental and idiopathic contexts. *Proceedings of the 8th International Seminar on Speech Production*. Universite March Bloch, Strasbourg.
- Shriberg LD, & Lof GL (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics*, 5(3), 225–279. doi:10.3109/02699209108986113
- Shriberg LD, Kent RD, McAllister T, & Preston JL (2019). *Clinical phonetics* (5th ed.). Pearson Publishing.
- Shriberg LD, Kwiatkowski J, & Hoffmann K (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research*, 27(3), 456–465. doi:10.1044/jshr.2703.456 [PubMed: 6482415]
- Shriberg LD, Strand EA, Fourakis M, Jakielski KJ, Hall SD, Karlsson HB, Mabie HI, McSweeny JL, Tilkens CM, & Wilson DL (2017). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: II. Validity studies of the pause marker. *Journal of Speech, Language, and Hearing Research*, 60(4), S1118–S1134. doi:10.1044/2016\_JSLHR-S-15-0297
- Shriberg L, Strand E, Jakielski K, & Mabie H (2019). Estimates of the prevalence of speech and motor disorders in persons with complex neurodevelopmental disorders. *Clinical Linguistics & Phonetics*, 33(8), 707–736. doi:10.1080/0s699206.2019.1595732 [PubMed: 31221012]
- Sjolander K, Beskow J (2019). *Wavesurfer* [Computer software]. <https://sourceforge.net/projects/wavesurfer/>
- Sparrow S, Cicchetti., & Balla D (2005). *Vineland adaptive behavior scales*. American Guidance Service.
- Steigler L (2015). Examining the echolalia literature: Where to do speech-language pathologists stand? *American Journal of Speech-Language Pathology*, 24(2), 750–762. doi:10.1044/2015\_AJSLP-14-016 [PubMed: 26161804]
- Stemberger JP, & Bernhardt BM (2020). Phonetic transcription for speech-language pathology in the 21st century. *Folia Phoniatica et Logopaedica*, 72(2), 75–83. doi:10.1159/000500701 [PubMed: 31550711]
- Stoel-Gammon C (1987). Phonological skills of 2-year-olds. *Language, Speech, and Hearing Services in Schools*, 18(4), 323–329. doi:10.1044/0161-1461.1804.323
- Stoel-Gammon C (1989). Prespeech and early speech development of two late talkers. *First Language*, 9(6), 207–223. doi:10.1177/014272378900900607

- Strand EA, Stoeckel R, & Baas B (2006). Treatment of severe childhood apraxia of speech: A treatment efficacy study. *Journal of Medical Speech-Language Pathology*, 14(4), 297–307.
- Strand E, & McCauley R (2019). *Dynamic evaluation of motor speech skill manual*. Brookes Publishing.
- 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. doi:10.1002/aur.1329 [PubMed: 24124067]
- Tager-Flusberg H, Plesa Skwerer D, Joseph RM, Brukilacchio B, Decker J, Eggleston B, Meyer S, & Yoder A (2017). Conducting research with minimally verbal participants with autism spectrum disorder. *Autism*, 21(7), 852–861. doi:10.1177/1362361316654605 [PubMed: 27354431]
- Tierney C, Mayes S, Lohs S, Black A, Gisin E, & Veglia M (2015). How valid is the checklist for autism spectrum disorder when a child has apraxia of speech? *Journal of Developmental and Behavioral Pediatrics*, 36(8), 569–574. doi:10.1097/DBP.000000000000189 [PubMed: 26114615]
- van Mourik M, Catsman-Berrepoets C, Paquier P, Yousef-Bak E, & van Dongen H (1997). Acquired childhood dysarthria: Review of its clinical presentation. *Pediatric Neurology*, 17(4), 299–307. doi:10.1016/s0887-8994(97)00081-7 [PubMed: 9436793]
- Velleman SL, & Strand K (1994). Developmental verbal dyspraxia. In Bernthal JE & Bankson NW (Eds.), *Child phonology: Characteristics, assessment, and intervention with special populations* (pp.110–139). Thieme Medical Publishers, Inc.
- Vitevitch M, & Luce P (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36(3), 481–487. doi:10.3758/bf03195594
- Vogel AP, & Maruff P (2008). Comparison of voice acquisition methodologies in speech research. *Behavior Research Methods*, 40(4), 982–987. doi:10.3758/BRM.40.4.982 [PubMed: 19001389]
- Vorperian H, & Kent R (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, 50(6), 1510–1545. doi:10.1044/1092-4388(2007/104)
- Yoder P, & Stone W (2006). A randomized comparison of the effect of two prelinguistic communication interventions on the acquisition of spoken communication in preschoolers with ASD. *Journal of Speech, Language, and Hearing Research*, 49(4), 698–711. doi:10.1044/1092-4388(2006/051)
- Yoder P, Watson L, & Lambert W (2015). Value-added predictors of expressive and receptive language growth in initially nonverbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 45(5), 1254–1270. doi:10.1007/s10803-014-2286-4 [PubMed: 25344152]
- Zhang C, Jepson K, Lohfink G, & Arvaniti A (2021). Comparing acoustic analyses of speech data collected remotely. *The Journal of the Acoustical Society of America*, 149(6), 3910–3916. doi:10.1121/10.0005132 [PubMed: 34241427]