The Stability and Validity of Automated Vocal Analysis in Preschoolers with Autism Spectrum Disorder in the Early Stages of Language Development

By

Tiffany Woynaroski

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Hearing and Speech Sciences

December, 2014

Nashville, Tennessee

Approved:

Chair: C. Melanie Schuele, Ph.D.

Co-Chair: Paul Yoder, Ph.D.

Stephen Camarata, Ph.D.

Bernard Rousseau, Ph.D.

To my five fabulous children, Logan, Bennett, Laney, Asher, and Lily,

and

To my husband, Adam, for his endless love and support

Acknowledgments

This work was funded by the National Institute for Deafness and other Communication Disorders (NIDCD R01 DC006893) and supported by the National Institute for Child Health and Disorders (NICHD) through the Vanderbilt Kennedy Center (P30HD15052) and a Preparation of Leadership Personnel grant (H325K080075; PI: Schuele), US Department of Education. The content is solely the responsibility of the author/s and does not necessarily represent the official views of the National Institutes of Health or the US Department of Education.

Table of Contents

List of Tables

Page

List of Figures

Page

List of Abbreviations

ADOS: Autism Diagnostic Observation Schedule

ASD: Autism spectrum disorders

CSBS-DP: The Communication and Symbolic Behavior Scales – Developmental Profile Behavior Sample

ICC: Intra-class correlation coefficient

LENA: Language Environment Analysis

LENA-DLP: Language Environment Analysis - Digital Language Processor

MB-CDI: MacArthur-Bates Communicative Development Inventory: Words and Gestures

PCS: Parent-child snack session

PCFP: Parent-child free play session

SSCS: Semi-structured communication sample with an examiner

SVI: Speech-related vocal island

Chapter 1

Introduction

Children with autism spectrum disorders (ASD) show a wide range of individual differences in their ability to use spoken words to communicate (Tager-Flusberg, Paul, & Lord, 2005). Children who can use a large number of spoken words to effectively and efficiently communicate can readily engage for social purposes across communication partners and settings. In contrast, those children who cannot use many spoken words to communicate struggle to convey messages that exceed the complexity of requesting solutions to their most basic wants and needs. Explaining individual differences in spoken word use of preschool children with ASD is especially important because learning to use words to communicate, or acquiring "useful speech," during the preschool years has been linked repeatedly with long-term outcomes in ASD (Billstedt, Carina Gillberg, & Gillberg, 2007; Eisenberg, 1956; Gillberg & Steffenburg, 1987; Kobayashi, Murata, & Yoshinaga, 1992; Lotter, 1974; Rutter, Greenfeld, & Lockyer, 1967).

A Transactional Approach to Explaining Individual Differences in Spoken Vocabulary in ASD

We take a transactional approach, which considers both child and parent factors, in explaining individual differences in spoken vocabulary use of preschoolers with ASD (McLean & Snyder-McLean, 1978; Sameroff & Chandler, 1975). In this study, we focus on one child factor, complexity of vocalizations, and one parent factor, linguistic input, that have been found to predict spoken language of preschoolers with ASD in studies involving conventional communication sampling methods (e.g., McDuffie & Yoder, 2010; Sheinkopf, Mundy, Oller, & Steffens, 2000; Siller & Sigman, 2002; Wetherby, Watt, Morgan, & Shumway, 2007) and a more novel approach - automated vocal analysis (Warren et al., 2010; Yoder, Oller, Richards, Gray, & Gilkerson, 2013). Our primary aim is to determine whether automated vocal analysis provides a valid and reliable alternative to conventional communication

sampling for measurement of these previously identified predictors of spoken language for the subset of preschoolers with ASD who are still in the early stages of language development.

Of the many previously identified predictors of spoken language for preschoolers with ASD, we chose to focus on child vocalization complexity and adult linguistic input because both theory and findings from previous research involving children with developmental delays suggest that there is a dynamic, or transactional, relationship between the complexity of children's vocalizations and parent linguistic input as children learn to use spoken words to communicate (e.g., Goldberg, 1977; Locke, 1996; Snyder-McLean, 1990; Woynaroski, Yoder, Fey, & Warren, in press). The earliest vocalizations produced by infants and young children with developmental delays are not very complex or "speech-like" relative to adult productions. However, over the course of development, children's vocalizations increase in complexity to include acoustic features that more closely resemble adult speech (Figure 1; Oller, 1995). Oller has called these acoustic parameters associated with more complex or adult-like speech, "infraphonological" features (Oller & Lynch, 1991).

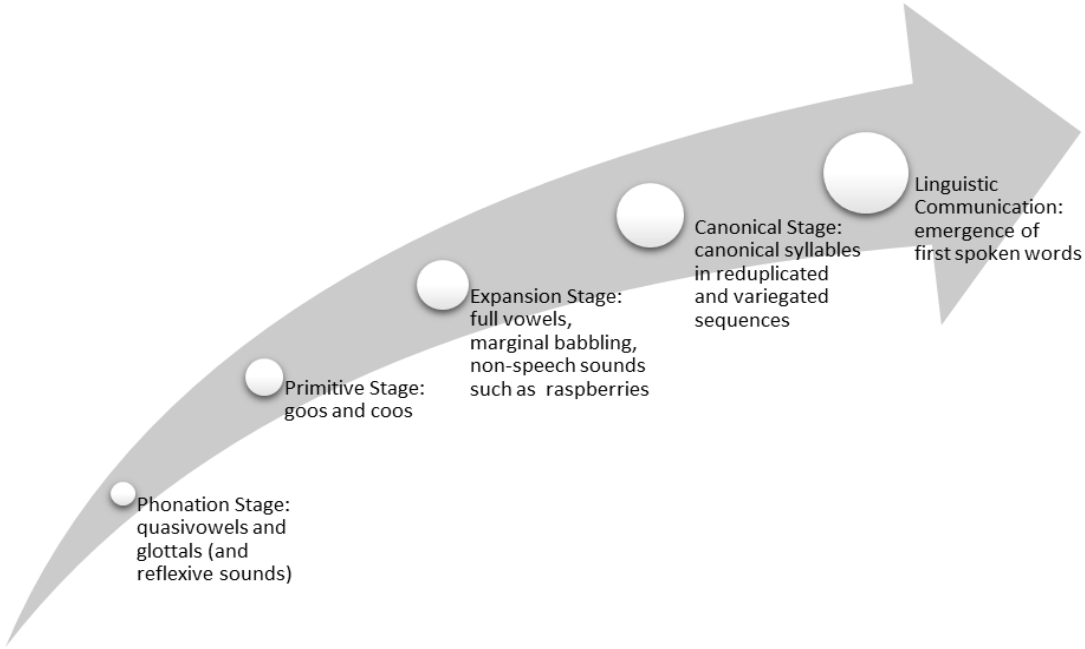


Figure 1. The stages of infraphonological development recognized by international consensus. Adapted from "Development of vocalizations in infancy," by D. K. Oller, 1995, In H. Winitz (Ed.), *Human communication and its disorders: A review* (Vol. IV), pp. 1-30. Copyright 1995 by York Press.

A transactional perspective presumes that the increase in child vocalization complexity over time, or infraphonological development, is influenced by adults in the child's environment in several different ways. For example, children may "tune in" to the features of the ambient language and attempt to produce sounds that are similar to the sounds produced by adults in the environment. Adults may then respond selectively to, or repeat, those child productions that are more speech-like or complex (Reeve, Reeve, Brown, Brown, & Poulson, 1992). These adult responses may reinforce the more speech-like productions of the child and provide phonological models that allow children to refine or "tune up" their vocalizations to more closely resemble those of the adult. As children produce vocalizations that are more complex, adults more frequently provide the spoken words that they think the child is trying to convey (Gros-Louis, West, Goldstein, & King, 2006), thereby scaffolding spoken vocabulary development.

A Need to Identify Nomologically Valid and Reliable Measures of Child Vocalization Complexity and Adult Linguistic Input

Given the theoretical basis and empirical support for child vocalization complexity and adult linguistic input as predictors of spoken vocabulary in children with ASD, there is a need to identify valid, or scientifically useful, measures of these constructs that may be used in research and applied in clinical practice. According to the nomological network theory of construct validation (Cronbach & Meehl, 1955), the construct validity of a measure may be expressed as a theoretically-based association between the measure and another construct of interest. In keeping with the nomological approach, we may then examine the construct validity of our candidate measures of child vocalization complexity and adult linguistic input by examining their correlations with spoken vocabulary of preschoolers with ASD. In this study, we are specifically interested in evaluating the nomological validity of automated vocal analysis relative to conventional communication sampling.

Importantly, the validity of any measure is limited by its reliability, particularly the type of reliability relating to the stability of estimates of a construct across observations or contexts (McCrae, Kurtz, Yamagata, & Terracciano, 2011). Stable measures result in a similar ranking of children on constructs of interest across observations. For example, when we are assessing the stability of child vocalization complexity as measured in conventional communication samples, we are interested in the extent to which we would make the same decisions regarding which children produce more complex or speech-like vocalizations and which children show less complex or speech-like vocalizations across the samples. We would conclude that conventional communication samples are yielding a fairly stable estimate of child vocalization complexity if we tend to rank children according to their vocal complexity in a very similar way in the first sample and in the second sample.

Mathematically, our ability to obtain a reliable or stable measure of a construct places an upper bound on validity. Within classical test theory, the predictive or concurrent validity for a measure cannot exceed the square root of the correlation between two versions of the same measure (Crocker & Algina, 1986). Practically, this simply means that our ability to explain the individual differences in spoken vocabulary of preschoolers with ASD with any measure is limited by our ability to obtain stable estimates of our putative predictors with that measure. Therefore, this study will additionally evaluate the stability of our measures of child vocalization complexity and adult linguistic input as derived via automated vocal analysis relative to indices of the same constructs as measured in conventional communication samples. Using Generalizability theory, which will be explained in the Methods section, we will estimate the number of samples across which scores need to be aggregated to attain stable estimates of our constructs of interest via automated vocal analysis and conventional communication sampling (Cronbach, 1972).

Child Vocalization Complexity As Measured in Conventional Communication Samples Predicts Spoken Language in Preschoolers with ASD

A few studies have previously demonstrated that child vocalization complexity as measured in conventional communication samples accounts for individual differences in spoken language of preschoolers with ASD. Wetherby et al. (2007) found that an index of consonant inventory from communicative child vocalizations predicted spoken language level, as well as nonverbal developmental level and communication symptom severity, approximately one year later in preschoolers with ASD. Scheinkopf and colleagues (2000) reported that the proportion of canonical syllables produced (out of total syllables produced) in vocalizations concurrently predicted expressive language age in a preverbal sample of preschoolers with ASD. In another study, Plumb and Wetherby (2013) found that a similar measure (i.e., the proportion of syllabic vocalizations out of total vocalizations) predicted both concurrent and future spoken language level in preschoolers with ASD.

Importantly, Plumb and Wetherby (2013) reported that the proportion of syllabic vocalizations *used communicatively* improved predictions of later spoken language above and beyond syllabic vocalizations produced for non-communicative purposes in their sample of young children with ASD. Previous studies have found that preschoolers with ASD differ from their typically developing peers in their communicative use of vocalizations in addition to the complexity of their vocalizations. For example, toddlers who are later diagnosed with ASD are less likely to vocalize when communicating and more likely to vocalize for a non-communicative purpose in comparison to typically developing peers of the same chronological age (Plumb & Wetherby, 2013; Shumway & Wetherby, 2009). One advantage of conventional communication sampling over the more novel approach, automated vocal analysis, may be the ability to tap into the complexity of vocalizations specifically produced for a communicative purpose. Our index of child vocalization complexity from conventional communication samples will be an aggregate of consonant inventory used communicatively as defined by Wetherby and colleagues (2007)

and an index of canonical syllable use in communication acts (Plumb & Wetherby, 2013; Sheinkopf et al., 2000). We explain why these were aggregated to produce one variable in the Methods section.

Adult Linguistic Input As Measured in Conventional Communication Samples Also Accounts for Individual Differences in Spoken Language Development in Preschoolers with ASD

Adult linguistic input as measured in conventional parent-child communication samples also predicts spoken language in preschoolers with ASD (e.g., Haebig, McDuffie, & Ellis Weismer, 2013a, 2013b; McDuffie & Yoder, 2010; Siller & Sigman, 2002, 2008). Parent linguistic input as measured in conventional communication samples has been conceptualized previously as comprising two broad categories: a) parent linguistic input that refers to and follows the child's attentional focus, and b) parent linguistic responses to child communication acts (McDuffie & Yoder, 2010). According to this classification system, parent linguistic input that follows into the child's attentional focus includes both parent utterances that follow into the child's focus of attention without redirecting the child's behavior (i.e., follow-in comments) and parent utterances that follow into the child's focus of attention and direct the child's behavior (i.e., follow-in directives). Parent linguistic responses to child communication acts include repetitions, expansions, or linguistic maps of the child's immediately preceding child communication act. Both parent linguistic input that follows into the child's attentional focus (e.g., Haebig et al., 2013a, 2013b; McDuffie & Yoder, 2010; Siller & Sigman, 2002, 2008) and parent linguistic responses to child communication acts (McDuffie & Yoder, 2010) have been found previously to predict spoken language of preschoolers with ASD.

It may be particularly important that adult linguistic input *follows the child's attentional or communicative lead* for preschoolers with ASD, who struggle to infer their communication partner's intentions and to shift their attentional focus to follow their communication partner's lead (Baird et al., 2000; Baron-Cohen et al., 1996; Zwaigenbaum et al., 2005). Although typically developing children are able to use their understanding of a communication partner's intentions and their ability to follow a

communication partner's focus of attention to learn words even when linguistic input does not follow the child's lead (Baldwin, 1993; Baron-Cohen, Baldwin, & Crowson, 1997; Dunham, Dunham, & Curwin, 1993; Tomasello & Farrar, 1986), children with ASD are less likely to learn words successfully when adults do not follow into the child's focus attentional focus (Baron-Cohen et al., 1997). Because automatic analysis cannot determine whether adult input refers to the referent of child attention or communication, measurement of adult linguistic input that specifically follows into the child's attentional and/or communicative lead may represent another advantage of conventional communication sampling over automated vocal analysis. Therefore, the index of adult linguistic input derived from our conventional samples will be parent linguistic responses to child attentional and communicative leads.

Automated Vocal Analysis as a Potential Alternative for Predicting Spoken Vocabulary of Preschoolers with ASD

Research spanning approximately the last decade thus provides some support for the nomological validity of conventional communication sampling procedures for measuring child vocalization complexity and adult linguistic input when the goal is predicting individual differences in spoken language use of preschoolers with ASD. However, the use of conventional communication sampling in research and the application of conventional communication sampling in everyday clinical practice are limited by a number of factors. First, it is unclear to what extent the behaviors displayed in brief communication samples collected in the laboratory are representative of behaviors that occur in everyday settings. Obtaining representative estimates for behaviors of interest often requires measurement of the behavior across several samples or contexts (Sandbank & Yoder, 2014; Yoder & Symons, 2010). Furthermore, coding of communication samples is limited by human perception. Previous studies have reported poor inter-rater reliability for some potentially important indices of child vocalization complexity, such as vocal quality, vowel duration, and atypical productions, and noted that

discrepancies were not always resolved due to problems with perception or objectivity (e.g., Plumb & Wetherby, 2013; Sheinkopf et al., 2000). Additionally, communication samples may produce unstable scores due to their brevity (Yoder & Symons, 2010). Finally, the large amount of time and high cost involved in collection and coding of even a few conventional communication samples is particularly prohibitive to application into clinical practice.

Emerging research suggests that a novel approach - automated vocal analysis - may provide a valid and reliable alternative for measurement of child vocalization complexity, as well as quantity of adult linguistic input, in preschoolers with ASD (Oller et al., 2010; Warren et al., 2010; Yoder et al., 2013). Automated vocal analysis allows for objective measurement of child vocalizations across a large number of infraphonological features and provides an estimate of adult linguistic input using day-long audio recorded samples collected in naturalistic settings. This method requires minimal time and cost (aside from the initial investment in hardware and software) relative to conventional communication sampling techniques, and thus may be more readily applied by researchers interested in measuring child vocalization complexity and adult linguistic input and ultimately by clinicians hoping to incorporate measurement of these important predictors in everyday practice.

Present Evidence for the Stability and Validity of Child Vocalization Complexity and Adult Linguistic Input as Measured Via Automated Vocal Analysis of Day-long Naturalistic Samples in ASD

One recent study found that a single day-long audio recording yielded a stable estimate of child vocalization complexity in preschoolers with ASD (Yoder et al., 2013). In the same study, the index of child vocalization complexity, based on 12 infraphonological features of speech-likeness (referred to as the vocal age equivalency score in Yoder et al., 2013), covaried with concurrent spoken vocabulary size and broader spoken language skill. In another investigation, an index of adult linguistic input (i.e., adult word count; Warren et al., 2010) as derived via automated vocal analysis was observed to correlate with concurrent spoken language level in preschoolers with ASD. The effect sizes observed in this work

suggest that variables derived via automated vocal analysis may even have a larger association with spoken vocabulary of preschoolers with ASD relative to variables measured in conventional communication samples (Warren et al., 2010; Yoder et al., 2013). For example, Yoder et al. (2013) observed a correlation of .73 between their selected index of vocal complexity from automated vocal analysis and the concurrent number of words that parents reported their child to say. In contrast, the mean correlation magnitudes observed for associations between indices of vocal complexity as measured in conventional communication samples and measures of spoken language is .49 (Plumb & Wetherby, 2013; Sheinkopf et al., 2000; Wetherby et al., 2007). Similarly, Warren et al. (2010) observed correlation coefficients ranging from .54 to .70 for associations between adult word count as derived via automated vocal analysis and various measures of spoken language skill in their sample. In contrast, the mean correlation magnitude observed for the relation between indices of adult linguistic input as measured in conventional communication samples and measures of spoken language is .45 (e.g., Haebig et al., 2013a, 2013b; McDuffie & Yoder, 2010; Siller & Sigman, 2002, 2008).

Gaps in the Current Literature

The aforementioned studies thus provide some support for the nomological validity and stability of automated vocal analysis in preschoolers with ASD. However, there is clearly a need for additional research into both the stability and the nomological validity of automated vocal analysis when the goal is predicting spoken vocabulary for the subset of preschoolers with ASD who are still in the earliest stages of language development. The samples in the aforementioned studies have largely comprised preschoolers with ASD who were already using words to communicate. As such, it is unclear whether a single day-long audio recording would yield a similarly stable estimate of child vocalization complexity in preschoolers with ASD who are still preverbal or who are just beginning to use words to communicate. One recent study found that it was necessary to average scores from multiple measurement contexts to obtain stable estimates of prelinguistic and linguistic skills in young children with intellectual disabilities,

particularly when the children were in the earliest stages of skill development (Sandbank & Yoder, 2014). Thus, we might suspect that several samples, whether audio recorded or more conventional in nature, may be necessary to derive stable estimates of child vocalization complexity in our sample of children with ASD who are in earlier stages of language development relative to the children who participated in the Yoder et al. (2013) study.

Furthermore, no previous investigation has examined whether a single day-long audio recording produces a stable estimate of adult linguistic input. However, we suspect that indices of adult linguistic input, regardless of how they are measured, may be unstable with a single sample because frequency measures of adult talk are likely to be highly contextually influenced (Yoder & Symons, 2010). Thus, the first objective of this study is to examine how many audio recorded samples are necessary to derive stable indices of child vocalization complexity and adult linguistic input relative to the number of conventional communication samples that are necessary to obtain similarly stable measures of these constructs in preschoolers with ASD who are in the early stages of language development.

Additionally, we do not know whether indices of child vocalization complexity and adult linguistic input as derived via automated vocal analysis correlate with concurrent spoken vocabulary in the subset of preschoolers with ASD who are largely still minimally verbal communicators. Even more importantly, we do not yet know whether these indices predict future spoken vocabulary in preschoolers with ASD who are still in the early stages of lexical development. Finally, we do not know the relative nomological validity of indices derived via automated vocal analysis versus indices measured in conventional communication samples as it applies to their associations with concurrent and future spoken vocabulary in this population. If variables derived via automated vocal analysis are “better than” or even “as good as” (i.e., non-significantly different from) variables measured in more conventional communication samples for predicting spoken vocabulary in this population, then an argument can be

made in favor of using the automated approach as an alternative to the more established communication sampling method in research and clinical practice.

Of course, automated vocal analyses, despite its many apparent advantages, may simply not be as valid as conventional communication sampling for predicting spoken vocabulary of children with ASD in the early stages of language development. The latter approach allows us to tap into aspects of child vocalization complexity (i.e., the complexity of vocalizations *used communicatively)* and adult linguistic input (i.e., parent linguistic input *that follows into the child's attentional and communicative lead)* that cannot be captured by the algorithms that drive automated vocal analyses. Thus, conventional communication sampling, though costly, time consuming, and subject to the limits of human perception, may measure child vocalization complexity and adult linguistic input in a way that allows us to more accurately explain individual differences in spoken vocabulary development in this subset of preschoolers with ASD. Therefore, our second objective is to determine whether indices of child vocalization complexity and adult linguistic input as derived via automated vocal analyses are superior to, or non-significantly different from, indices of the same constructs as measured in conventional communication samples for predicting concurrent and future spoken vocabulary in preschoolers with ASD in the early stages of spoken language acquisition.

Research Questions

Our specific research questions are:

a) How many day-long audio recording sessions are required to derive stable measures of child vocalization complexity and adult linguistic input in a sample of preschoolers with ASD in the early stages of language development?
b) How many conventional communication samples are required to derive stable measures of child vocalization complexity and adult linguistic input in our sample of preschoolers with ASD in the early stages of language development?

c) Is child vocalization complexity as measured via automated vocal analysis “better than” or at least “as good as” child vocalization complexity as measured in conventional communication samples for predicting concurrent and future spoken vocabulary in preschool children with ASD in the early stages of language development?

d) Is adult linguistic input as measured via automated vocal analysis “better than” or at least “as good as” parent linguistic responses as measured in conventional communication samples in predicting concurrent and future spoken vocabulary in preschool children with ASD in the early stages of language development?

Chapter 2

Method

Overview of Study Design

To answer these research questions, this study drew on a subset of extant data from a large-scale, longitudinal correlational investigation that sought to identify predictors of spoken language in young children with ASD who were preverbal or minimally verbal at entry to the study. For all participants in the larger longitudinal study, spoken vocabulary was assessed by parent report at five time points: a) at entry to the study, b) 4 months after entry to the study, c) 8 months after entry to the study, d) 12 months after entry to the study, and e) at a final assessment period 16 months after entry to the study. At each of these time points, child vocalization complexity was assessed via two conventional communication samples involving the child and examiner. These communication samples were coded to derive an aggregate measure of consonant inventory in communication acts as previously defined by Wetherby and colleagues (2007) and proportion of communication acts with canonical syllables. Conventional parent-child communication samples were also collected at two time points for the larger longitudinal study: a) 4 months after entry to the study, and b) 12 months after entry to the study. These samples were coded to derive our proposed index of adult linguistic input from conventional communication samples, parent linguistic responses to child attentional and communicative leads.

A subset of children recruited for the larger longitudinal investigation also participated in this supplemental study, in which complexity of child vocalizations and quantity of adult linguistic input was measured via two day-long audio recordings collected in the child's natural settings and analyzed via automated vocal analysis. Children participated in the supplemental study at various time points relative to the larger longitudinal study. That is, day-long audio recordings were collected at entry to the longitudinal study for some children, but at later measurement periods for others based solely on the

time at which funding for the supplement was obtained. Thus, the putative predictors of interest for the present study included: a) child vocalization complexity and adult linguistic input as derived via automated vocal analysis at the time of the two day-long audio recordings (Time 1 for the present study), and b) child vocalization complexity and adult linguistic input as measured in conventional communication samples at the period for the larger longitudinal study that was most temporally proximal to the collection of day-long audio recordings (Time 1 for the present study). Outcomes of interest for the present study included: a) concurrent child spoken vocabulary at the period for the larger longitudinal study that was most temporally proximal to the collection of day-long audio recordings (Time 1 for the present study), and b) future child spoken vocabulary 4 months later (Time 2 for the present study).

Recruitment and study procedures were conducted in accordance with the approval of the Vanderbilt University Institutional Review Board. Parents signed a written informed consent prior to participation in the study. All families were compensated for their participation.

Participants

Participants included 33 preschool children with ASD (29 male; 4 female) who were recruited at Vanderbilt University in Nashville, TN. Children included in this supplemental study were required to meet the following inclusion criteria at the time of their day-long audio recordings: a) chronological age between 24 and 48 months; b) diagnosis of an ASD based on Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000) and clinical judgment of a licensed psychologist according to criteria from the Diagnostic and Statistical Manual of Mental Disorders – 4th Edition (DSM-IV; American Psychiatric Association, 2000); c) no severe sensory or motor impairments; d) no identified metabolic, genetic, or progressive neurological disorders; e) reported spoken vocabulary of less than or equal to 200 words on the MacArthur-Bates Communicative Development Inventory: Words and Gestures (MB-CDI; Fenson et al., 2003) vocabulary checklist (i.e., the approximate vocabulary size at which most children would be

expected to be combining words according to Bates, Dale, & Thal, 1995); and f) primarily English-speaking household. All participants in this supplemental study were diagnosed with Autistic Disorder based on ADOS total algorithm score and clinical impression of the evaluating psychologist. Additional sample characteristics for child participants at Time 1 for the present study are summarized in Table 1.

Table 1

Sample Characteristics

Characteristic	Mean	SD	Minimum	Maximum
Chronological age in months	39.67	6.50	25.58	48.00
Mullen composite standard score	51.13	4.94	49	68
Number of words spoken on MB-CDI	26.18	42.04	0	156
ADOS algorithm total score	23.70	3.55	16	28

Note. Mullen = Mullen Scales of Early Learning (Mullen, 1995); MB-CDI = MacArthur Bates Communicative Development Inventory: Words and Gestures (Fenson et al., 2003); ADOS = Autism Diagnostic Observation Schedule (Lord et al., 2000).

The parent who served as the primary caregiver for each participant self-reported the highest level of formal education that they had achieved by checking one of nine educational levels on a demographic questionnaire. Parents reported achieving a mean formal education level of 1-2 years of college or technical school. Parent reported level of formal education achieved is further summarized in Table 2.

Table 2

Summary of Primary Caregiver Formal Education Level

Elementary through High School	
0 years	1 (3.03%)
1-6 years	0 (0%)
7-9 years	0 (0%)
10-11 years	2 (6.06%)
12 years or GED	6 (18.18%)
College or Technical School	
1-2 years	9 (27.27%)
3-4 years	9 (27.27%)
Graduate or Professional School	
1-2 years	3 (9.09%)
3-4 years	3 (9.09%)

Data Collection

Conventional measures of child vocalization complexity and adult linguistic input. Child vocalization complexity was measured via two conventional communication samples with an examiner: a) the Communication and Symbolic Behavior Scales - Developmental Profile Behavior Sample (CSBS-DP; Wetherby & Prizant, 2002), and b) a semi-structured communication sample with an examiner (SSCS; Yoder & Stone, 2006). Parent linguistic responses to child communicative and attentional leads were measured via two conventional parent-child communication samples: a) a parent-child free play session (PCFP), and b) a parent-child snack session (PCS).

CSBS-DP. The CSBS-DP is a standardized, structured communication sample conducted by an examiner who is unfamiliar to the child outside the assessment context. The child's primary caregiver is present throughout the sample. The caregiver is instructed to respond naturally to the child, but not to model novel behaviors or to direct the child's behaviors. The CSBS-DP comprises: a) a series of communicative temptations, during which the child is presented with a wind-up toy, bubbles, a balloon, a jar with food, a bag with toys, and a set of board books; b) a symbolic play segment with a stuffed animal and dishes set; c) two gaze and point following probes; d) comprehension questions for person names, body parts, and object names; and e) and a constructive play component with building blocks. The total duration of the CSBS-DP is approximately 20-30 minutes. This communication sampling context has been used in previous studies reporting associations between child vocalization complexity (i.e., proportion of canonical syllables produced and number of different consonants produced in communicative vocalizations) and spoken language in preschoolers with ASD (e.g., Sheinkopf et al., 2000; Wetherby et al., 2007). This sample was administered according to the CSBS-DP manual (Wetherby & Prizant, 2002).

SSCS. The SSCS is a slightly less structured sampling context relative to the CSBS-DP, but it is also conducted by an examiner who is unfamiliar to the child outside the assessment context. The examiner allows the child to select from a standard set of developmentally appropriate toys (cause and effect pop-up toys, race cars, plastic people and animals, a farm puzzle, a shape sorter, stack and roll cups, building blocks, and a xylophone), according to the child's degree of interest and engagement with each toy. The examiner follows the child's attentional lead and joins in the child's play activities. The examiner provides limited scaffolding for communication and play behaviors, serving primarily as a respondent to child initiations. Direct prompts for child communication and actions are not provided unless the child is disengaged for a prolonged period of time (i.e., more than 10 seconds). The total duration of the SSCS is 15 minutes. This procedure was used in addition to the CSBS-DP to obtain a

more stable estimate of child communication behaviors (see the Variables section for additional information regarding aggregation across measures; Sandbank & Yoder, 2014).

PCFP. The PCFP sampling context is an unstructured communication sample involving the child and his/her primary caregiver. The PCFP involves 10 minutes of free play and 5 minutes of unstructured book sharing between each parent-child dyad. In the free play component, parents are provided with two standard sets of toys: a) stacking/nesting buckets, a baby doll and bottle, a ball, a rattle, plastic snap beads, a cause and effect pop-up toy, a small car, and a set of colored bead strands; and b) a Fisher-Price farm set with accessories and a slinky. Parents are instructed that they should offer their child a choice of the two play sets and proceed to play as they would typically play at home. They are also told that they can switch to the second toy set if their child becomes disinterested in the first set. Following the free play, the examiner provides the parent with three board books. Parents are told that they should "look at" the books with their child for 5 minutes, but they are not directly instructed to read the books. Parents are not directed to respond to child communication acts in any particular way. Similar parent-child interactions have been used in previous studies finding links between parent linguistic responses to child leads and later spoken language in preschoolers with ASD (e.g., Haebig et al., 2013b; McDuffie & Yoder, 2010; Siller & Sigman, 2008).

PCS. The PCS is an unstructured parent-child interaction that revolves around a familiar routine – snack time. The snack context was selected because it provides numerous opportunities for child communication, particularly requests, and for parent linguistic responses. The parent and child are seated at a table and chairs and are provided with a standard set of items including: two plates, two cups with removable lids, two spoons, a snack in a clear container with a screw-on top, and a drink in a clear pitcher. Snack and drink items are provided or approved by the parent. The parent is instructed to interact with the child as they would during a typical snack time at home. The total duration of this communication sample is 10 minutes. This procedure was used in addition to the PCFP to increase

sampling of parent behaviors (see the Variables section for additional information regarding aggregation across measures; Sandbank & Yoder, 2014).

Measure of child spoken vocabulary. Spoken vocabulary was measured via the MB-CDI vocabulary checklist. Parents were asked to check items on the MB-CDI to indicate whether their child either "understands" or "understands and says" early lexical items in categories such as actions, household items, and animals. MB-CDIs were mailed to each participant's home prior to assessment visits. Parents completed the checklists at their convenience and reviewed their responses with a member of the research team in the clinic during their scheduled assessment visit (or by phone in those few instances that participants did not return to the clinic for additional assessments).

Audio recordings of child vocalizations and adult linguistic input. Child vocalization complexity and adult linguistic input were additionally assessed via two day-long audio recordings collected on consecutive days in the child's natural settings. Prior to their recording period, parents of all participants received an audio recording packet that included fully-charged, compact, and light-weight digital recorders called Language Environment Analysis - Digital Language Processors (LENA-DLPs; LENA Research Foundation, 2014), a specialized vest with a built-in pocket for the recording device, and step-by-step recording instructions. Parents were instructed to turn on the LENA-DLP when their child woke up in the morning, place the recorder in the front pocket of the specially designed clothing, and allow the recorder to run continuously for a full 16-hour day (i.e., max LENA-DLP recording time) on two consecutive days. Parents were directed to remove the specialized clothing during bathing and sleeping periods, but to place it near the child to continue recording the language environment. Parents returned materials to the research team by mail at no cost once both audio recordings had been collected. Upon receipt of participant materials, research staff transferred audio files from the digital recorders directly to a computer for analysis.

Data Reduction

Coding of conventional communication samples. The four conventional communication samples were coded using ProcoderDV software (Tapp & Walden, 1993) by personnel who completed a training protocol including: a) review of the coding manuals, b) practice coding communication samples, and c) coding of three consecutive samples with 80% accuracy or higher for each communication sampling context. Each communication sample was coded by a primary coder, and 20% of the samples were randomly selected for independent coding by a secondary coder. To prevent observer drift, a coding supervisor reviewed interval-by-interval discrepancies between primary and secondary code files to identify potential sources of disagreement and to resolve any discrepancies. The coding supervisor also regularly corresponded with both primary and secondary coders regarding any concerns.

The two conventional communication samples with an examiner, the CSBS-DP and SSCS, were coded for intentional child communication acts using a five second partial interval coding system. Intentional child communication acts were defined as: a) vocal or gestural acts combined with coordinated attention to object and person; b) conventional gestures (e.g., showing, pointing) with attention to an adult; and c) symbolic forms (i.e., words, sign language). Intervals coded for child communication acts were subsequently coded for the production of canonical syllables and the number of different consonants used within communication acts. Canonical syllables were defined as vocalizations in which a rapid transition occurred between vowel-like and consonant-like speech sounds. The number of different consonants used communicatively were coded according to Wetherby's CSBS-DP True Consonant Inventory List, which includes supraglottal consonants that emerge earliest or are produced relatively frequently by young children and that are easy to code (Wetherby & Prizant, 2002).

The two conventional parent-child samples, the PCFP and PCS, were coded for child attentional leads using a five second partial interval coding system. Child attentional leads were defined as looking at a referent (e.g., parent, snack item, toy item, chair) or actively touching a referent for at least one

second. The PCS was also coded for child communicative leads using a five second partial interval coding system. Child communicative leads in the PCS were coded as: a) vocal or gestural acts combined with coordinated attention to object and person; b) conventional gestures (e.g., showing, pointing) with attention to adult; and c) symbolic forms (i.e., words, sign language). It was determined a priori that child communicative leads could not be coded for the PCFP sample. In this sample, parents are simply instructed to play with their child as they would play at home. Parents are not directed to position themselves or their children in any specific manner relative to the camera throughout the sample because such instructions may impact the parent-child interaction (e.g., by causing the parent to restrict their own behaviors or to adopt a more directive/re-directive style with their child). Unfortunately, the free parent-child positioning prevents us from maintaining optimal camera angles necessary to code child communicative leads.

Intervals coded for a child attentional or communicative lead within the PCS and PCFP were subsequently coded for parent linguistic response. Parent linguistic responses were coded following child attentional leads when the parent produced an utterance that: a) was about an object, person, or event within the child's present or immediately preceding focus of attention; and b) conveyed a meaning relevant to the child's focus of attention (i.e., about the object, person, or event, or about the properties or qualities of the object, person, or event). Parent linguistic responses to child attentional leads included utterances that did and did not direct the child's behavior (i.e., utterance types that have been referred to as follow-in directives and follow-in comments in previous research). Parent linguistic responses to child communicative leads: a) put into words the presumed meaning of the child's immediately preceding nonverbal intentional communication act; b) involved a repetition of the child's word approximation using adult pronunciation; or c) expanded (i.e., added words to) the child's utterance attempt. These types of parent linguistic responses have been referred to as linguistic maps, repetitions, and expansions in previous work. Although it is possible to distinguish between parent

linguistic responses to child attentional and communicative leads, we did not do so because both types of responses are theoretically and empirically linked to later child language through the same associational mechanism (McDuffie & Yoder, 2010).

Automated vocal analysis. Automated vocal analysis of day-long audio recordings was carried out using LENA software and analytic software developed by Oller and colleagues (2010). LENA software uses modified speech recognition algorithms to segment the audio stream and to classify segments according to their statistical similarity to one of nine different pre-defined sound sources: target child, other child, adult male, adult female, overlapping speech/sound, electronic media, noise, silence, or unclear. Segments identified as being produced by the target child are classified as speech-related child utterances (i.e., a speech-related vocalization of at least 50 ms duration bounded by sounds of other source types for more than 300 ms), fixed signals (such as cries or screams), or vegetative sounds (such as burps). Speech-related child utterances may comprise one or more speech-related vocal islands, defined as 50 ms or greater of target child-produced, speech-related sounds with high acoustic energy bounded by periods of low acoustic energy. Speech-related vocal islands are intended to correspond with "syllable-like" segments of the speech-related child utterance (though it should be noted that "speech-related vocal island" is not synonymous with the linguistic concept of "syllable").

Child speech-related vocal islands were analyzed for complexity or "speech-likeness" in accordance with the procedures of Oller and colleagues (2010). Speech-related vocal islands were assessed for 12 acoustic features in four perceptual categories (Table 3) that are supported by both infrastructural vocal theory and principal component analyses carried out in a previous normative study (Oller, 2000; Oller et al., 2010). Each speech-related vocal island is automatically scored for the presence or absence of each parameter based on criterion values, and each of the 12 parameters receives a raw score for speech-likeness based on the proportion of speech-related child utterances with at least one vocal island demonstrating the criterion level for the acoustic property of interest.

Table 3

The 12 Acoustic Parameters Used in Automated Analysis of Child Speech-Related Vocal Islands

	Rhythm/Syllabicity	
1	Voiced: Pitch Detectable for 50% SVI*	Positive classification on these parameters suggests that vocalizations tended to show voicing features, canonical formant transitions, and spectral entropy variations consistent with speech-like rhythm and syllabicity.
2	Canonical Syllable: Formant Transitions < 120 ms*	
3	Spectral Entropy Typical of Speech*	
	Low Spectral Tilt and High Pitch Control	
4	Mean Pitch High (Squeal): > 600 Hz*	Positive classification on these parameters suggests more active emotional expression in the high spectral frequency range (i.e., a squeal quality to vocal productions).
5	Low Spectral Tilt*	
6	High Frequency Energy Concentration	
	Wide Format Bandwidth and Low Pitch Control	
7	Mean Pitch Low (Growl): < 250 Hz	Positive classification on these parameters suggests more active emotional expression in the low spectral frequency range (i.e., growl quality to vocal productions).
8	Wide Bandwidth (First Two Formants)*	
	Duration of SVIs within Utterances	
9	Short (110-250ms)*	Positive classification on parameters 9 and 10 suggests speech-like rhythmic organization because duration values are typical of syllables in speech.
10	Medium (250-600ms)*	
11	Long (600-900ms)	Positive classification on parameters 11 and 12 suggest the opposite because the corresponding durations are beyond the range of typical syllables.
12	Extra Long (900-3000ms)*	

Note. SVI = speech-related vocal island. Table adapted from "Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development," by Oller, D. K., Niyogi, P., Richards, J.A., Gilkerson, J., Xu, D., Yapanel, U., & Warren, S.F., 2010, *PNAS, 107*, p. S24. Copyright [2010] by the National Academy of Sciences. Adapted with permission.
* indicates that ASD and TD groups significantly differed in mean normalized values for this parameter in Oller et al., 2010.

The index of child vocalization complexity as measured via automated vocal analysis that will be utilized in the present study is the infraphonological vocal complexity score. The infraphonological vocal complexity score is an aggregate of the 12 parameter raw scores, weighted by the unstandardized regression coefficients from a multiple regression equation that predicted chronological age based on the 12 parameter scores in the normative sample of Oller et al. (2010). Previous studies have demonstrated that this score (previously called the vocal development age equivalency score by Yoder and colleagues and unnamed in the original study by Oller et al.) indexes vocal development in typically developing preschoolers (accounts for 69% of the variance in chronological age), differentiates preschool children with ASD from typically developing controls, and accounts for individual differences in concurrent spoken language in preschoolers with ASD (Oller et al., 2010; Yoder et al., 2013).

The LENA software further analyzes audio segments identified as adult male and adult female speech to generate a score for adult word count, which is an estimate of proximal (i.e., within approximately 4-6 feet) adult linguistic input. The processing algorithm estimates the number of adult words in each adult speech segment based on acoustic properties of the speech signal (e.g., algorithm-identified consonant and vowel counts, total duration of adult production). Although the software provides only an estimate of adult word count, previous work has shown that the estimates derived via automated vocal analysis of day-long audio recordings correspond highly with adult word counts of human transcribers (i.e., with automated vocal analysis adult word count estimates corresponding with human transcriber word counts at approximately 98%; Xu, Yapanel, & Gray, 2008). Warren et al. (2010) have previously found that this particular index of adult linguistic input correlates with concurrent spoken language in preschoolers with ASD. Further information about LENA software processing specifics, as well as the validity and reliability of the automated measures, is available in previous reports (Xu, Yapanel, & Gray, 2008; Xu, Yapanel, Gray, & Baer, 2008).

Variables. To increase the likelihood of obtaining a stable, and thus valid, estimate for our putative predictors, we aggregated across multiple measurement contexts or sampling days and/or conceptually similar and empirically related component variables in deriving variables when warranted (Rushton, Brainerd, & Pressley, 1983; Sandbank & Yoder, 2014; Tager-Flusberg et al., 2009). Our criterion level for evidence of an empirical relation amongst component variables with content validity for measuring the same construct (i.e., child vocalization complexity or adult linguistic input) was a minimum covariation of .40 prior to aggregation (Cohen & Cohen, 1984). We derived aggregates by averaging z-scores for component variables of different scales and for samples of different durations. See Table 4 for a summary of constructs, measurements, and metrics according to time period and research question.

Table 4

Constructs, Measurement Procedures, Variables, Periods, Roles (and Corresponding Research Questions)

Construct	Measures	Variable metric(s)	Period	Role (RQ)
Spoken Vocabulary	MB-CDI	Number of words that the child is reported to say on the MB-CDI[†]	1,2	Dependent variable (c, d)
Child vocalization complexity derived via automated vocal analysis	LENA-DLP recorded day-long sample	Infraphonological vocal complexity score	1 1	Dependent variable (a) Predictor (c)
Child vocalization complexity measured in conventional communication samples	CSBS-DP SSCS	Proportion of communication acts including canonical syllables aggregated with number of different consonants used communicatively	1	Dependent variable (b) Predictor (c)
Adult linguistic input derived via automated vocal analysis	LENA-DLP recorded day-long sample	Adult “word” count per hour aggregated across audio recordings	1 1	Dependent variable (a) Predictor (d)
Adult linguistic input measured in conventional communication samples	PCS PCFP	Parent linguistic responses to child attentional or communicative leads	1	Dependent variable (b) Predictor (d)

Note. CSBS-DP = Communication and Symbolic Behavior Scales - Developmental Profile Behavior Sample (CSBS-DP; Wetherby & Prizant, 2002); LENA-DLP = Language Environment Analysis – Digital Language Processor; MB-CDI = MacArthur Bates Communicative Development Inventory: Words and Gestures vocabulary checklist (Fenson et al., 2003); PCS = parent-child snack session; PCFP = parent-child free play sample, SSCS = semi-structured communication sample with the examiner. Time 1 indicates the time period from the larger longitudinal study most proximal to each child's audio recordings. Time 2 indicates the time period from the larger longitudinal study subsequent to the period most temporally proximal to audio recordings.
[†] Time 1 and 2 MB-CDI scores were log10 transformed to correct severe positive skew.

Child vocalization complexity as measured in conventional communication samples. To derive the variable for child vocalization complexity as measured in conventional communication samples, we aggregated across two component variables that were derived by concatenating across the CSBS-DP and SSCS samples: a) the proportion of communication acts including canonical syllables, and b) the number of different consonants from Wetherby's True Consonant Inventory List (Wetherby & Prizant, 2002) used communicatively.

Adult linguistic input as measured in conventional communication samples. Our variable for adult linguistic input as measured in conventional communication samples was the number of parent linguistic responses to child attentional and communicative leads across the PCFP and PCS communication sampling contexts.

Child vocalization complexity as derived via automated vocal analysis. At the outset of this study, it was not clear whether a single day-long audio recording would produce a stable estimate of vocal complexity in preschool children with ASD in the early stages of language development. We planned to use the infraphonological vocal complexity score from the first 16 hour audio recording as our index of child vocalization complexity as measured via automated vocal analysis if our analyses indicated that a single day-long recording was sufficient to obtain a stable estimate of vocalization complexity in our sample, but to average across the two consecutive day-long audio recordings if a single day-long audio recording was not found to yield a reliable estimate for the infraphonological vocal complexity score in our sample.

Adult linguistic input as measured via automated vocal analysis. No previous studies had examined the stability of our index of adult linguistic input as derived via automated vocal analysis, adult word count. Thus, it also was not clear whether a single day-long recording would yield a reliable estimate of adult linguistic input. Following the same logic as indicated for the vocal complexity variable from automated analysis, we planned to aggregate this score across the two day-long audio recordings obtained for each participant at Time 1 if a single day's audio recording was insufficiently stable. Prior to aggregation, we prorated (i.e., divided by time observed) this variable to adult word count per hour to account for variability in the length of participants' audio recordings.

Spoken vocabulary. The dependent variables for concurrent spoken vocabulary and future spoken vocabulary were the raw number of words the child was reported to say on the MB-CDI at Time 1 and Time 2, respectively.

Data Analysis

Imputation of missing data. When variables with missing data were correlated with other variables with existing data, we planned to impute missing data points using a state-of-the-art approach to missing data analyses – multiple imputation (Enders, 2011b). Briefly, multiple imputation involves generation of multiple data sets (e.g., 20) with plausible values for missing data points, analysis of each filled-in data set, and pooling of the information from the multiple data sets into a single result. This method is preferable to traditional methods for dealing with missing data (e.g., deletion, single imputation, last observation carried forward) in longitudinal data sets because it prevents loss of information related to missing data, reduces bias, improves parameter estimates, and preserves statistical power to detect effects of interest (Enders, 2011a). Incorporation of auxiliary variables, which are ancillary to the primary research questions, but potential correlates of missingness or of the missing variable, during the imputation phase reduces bias and improves power by infusing the dataset with some of the information that would otherwise be lost in the missing data (Collins, Schafer, & Kam, 2001;

Enders, 2011b). Our data set included many variables that were likely correlates of our predictor and/or outcome variables. For example, parent level of formal education was a potential auxiliary variable for imputation of missing data for adult linguistic input as measured in conventional communication samples or via automated vocal analysis. Simulations suggest that auxiliary variables are most useful when their correlation with the missing variable is > .40 (Enders, 2011b). Therefore, we sought to identify auxiliary variables that showed associations of approximately this magnitude or greater for each variable with missing data points.

Evaluation of the stability of variables derived via automated vocal analysis and conventional communication samples. Four Generalizability and Decision (G & D) studies, one per predictor, were conducted to determine how many day-long audio recordings and conventional communication samples are necessary to derive stable measures of child vocalization complexity and adult linguistic input in our sample of preschoolers with ASD in the early stages of language development (Research Questions a and b). G studies allow us to use the variance estimates derived from an analysis of variance to estimate the proportion of variance for each variable that is attributable to: a) what we intend to measure (i.e., individuals differences in child vocalization complexity or adult linguistic input among participants), and b) how we are measuring the construct of interest (i.e., day-long naturalistic audio recorded samples, conventional communication samples) (Shavelson & Webb, 1991).

In each G study, we used either the two consecutive day-long audio recordings or the two conventional communication samples acquired for each participant at Time 1. The total variance in the reliability sample included variance attributable to individual differences among participants plus error variance. We estimated the proportion of overall variance that was attributable to individual differences among participants by using the average across audio recording day or across communication sampling context as an estimate of the participants' "true scores" for each variable. We estimated the proportion of variance due to measurement error by adding the variance due to differences between audio

recording days or communication sampling contexts and the variance due to the interaction between participants and audio recording day or communication sample. Our estimate of stability, a type of intra-class correlation coefficient (ICC) referred to as a *g* coefficient, was derived by dividing the proportion of variance attributable to the individual differences among participants by the total variance.

In the D studies, we used the data derived from the G studies to determine the number of day-long audio recordings or communication samples across which we would need to average to yield stable estimates of child vocalization complexity and adult linguistic input. The D studies allowed us to project ICCs for our indices of child vocalization complexity and adult linguistic input, beyond the number of day-long audio recordings actually collected or communication samples actually observed using a similar logic to the Spearman prophesy formula (Yoder & Symons, 2010). The a priori criterion that we selected as indicative of "short term stability" for each dependent variable was an ICC of .80. This is a very conservative or "strict" criterion for stability (Bakeman, McArthur, Quera, & Robinson, 1997), but it is consistent with a commonly used benchmark for "good" reliability in the field of Special Education (Horner et al., 2005).

Evaluation of the validity of variables derived via automated vocal analysis and conventional communication samples. Pearson's product-moment correlation coefficients were obtained to explore the extent to which each of our predictors positively correlated with concurrent (Time 1) and future (Time 2) spoken vocabulary in our sample without controlling for any other factors (i.e., the magnitude of zero-order correlations). The a priori alpha level established for statistical significance was $p < .05$. Tests were one-tailed because theory and extant data suggested that all of these correlations should be positive.

We subsequently used Steiger's Z-test to directly compare the magnitude of the correlations for our putative predictors with spoken vocabulary outcomes (Lee & Preacher, 2013; Steiger, 1980).

Steiger's Z tests the difference between two dependent correlations with one variable in common by converting each correlation into a z-score using Fisher's r to z transformation and using Steiger's equations to compute the asymptotic covariance of the estimates. These values are then compared using an asymptotic z-test. Steiger's Z tests were carried out using an online utility developed by Lee & Preacher (2013).

Power analysis. Power analyses conducted with the G*Power 3 program (Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007) with consideration of the observed effect sizes for associations for indices of both child vocalization complexity and adult linguistic input as measured in conventional communication samples or derived via automated vocal analyses with spoken language in preschool children with ASD indicated that this study was powered at greater than .8 to detect effects of interest with 32 participants (Plumb & Wetherby, 2013; Sheinkopf et al., 2000; Warren et al., 2010; Wetherby et al., 2007; Yoder et al., 2013).

Chapter 3

Results

Preliminary Analyses

Interobserver reliability. Interobserver reliability was estimated using absolute agreement intra-class correlation coefficients (ICCs). When used to estimate interobserver reliability, ICCs reflect the proportion of variance that is due to among participant variability in true score estimates for the behaviors of interest versus measurement error that is due to inconsistency between coders. In group designs, it is desirable for the index of reliability to reflect both of these types of information, rather than just reflecting the session-by-session agreement between coders. The inclusion of variance among participants in the ICC is important because our ability to show a relation varies according to the degree of true score variance among participants. When the true score variance among participants is small, interobserver agreement must be extremely high (i.e., approach 1.0) to detect the true differences among participants. Conversely, when the true score variance among participants is large, even a low small/large agreement is sufficient to detect a relation (Yoder & Symons, 2010). ICCs range from -1.0 – 1.0 with values greater than .7 indicating a high degree of consistency or "very good" agreement in ranking of participants on the behaviors of interest across coders (Yoder & Symons, 2010). The mean ICC for canonical syllabic communication across CSBS and SSCS samples, calculated in a way that included both unitizing errors (i.e., errors in identifying the presence of a communication act) and classifying errors (i.e., errors in identifying whether the act has a canonical syllable), was .96 (range .94 - .97). The mean ICC for consonant inventory as aggregated across CSBS-DP and SSCS samples was .96 (range .93 - .98). The mean ICC for parent linguistic responses to child attentional and communicative leads in the PCFP and PCS samples was .95 (range .90 – .99).

Aggregation of component variables from conventional communication samples. The intercorrelation among the proportion of communication acts including canonical syllables and the

number of different consonants from Wetherby's True Consonant Inventory List (Wetherby & Prizant, 2002) used communicatively across the CSBS-DP and SSCS samples surpassed the .40 threshold for aggregation as an index of child vocalization complexity as measured in conventional communication samples ($r(31) = .84$, $p < .001$, one-tailed). The intercorrelation among parent linguistic responses to child attentional leads from the PCFP and parent linguistic responses to child attentional and communicative leads in the PCS also surpassed this empirical criterion for aggregation as an index of adult linguistic input as measured in conventional communication samples ($r(25) = .57$, $p = .002$, one-tailed).

Transformation of variables. The analysis method that we used in evaluating the validity of our measures of child vocalization complexity and adult linguistic input assumes multivariate normality (Enders, 2011b). Multivariate normality is more likely when univariate distributions do not grossly depart from the normal distribution (Tabachnick & Fidell, 2001). All predictors and dependent variables were evaluated for normality. Variables showing univariate skewness > |.8| or kurtosis > |3.0| were transformed prior to imputation and analysis. MB-CDI scores were positively skewed at both Time 1 and Time 2, but were corrected with log10 transformation.

How Many Audio Recordings are Required to Derive Stable Estimates of Child Vocalization Complexity and Adult Linguistic Input?

The infraphonological vocal complexity score surpassed our .8 threshold for acceptable stability with one day-long audio recording (g = 0.82) and showed only slight increases in stability with additional samples. This result confirms that a single audio recording is sufficient to obtain a stable estimate of child vocalization complexity in preschool children with ASD in the early stages of language development (Figure 2).

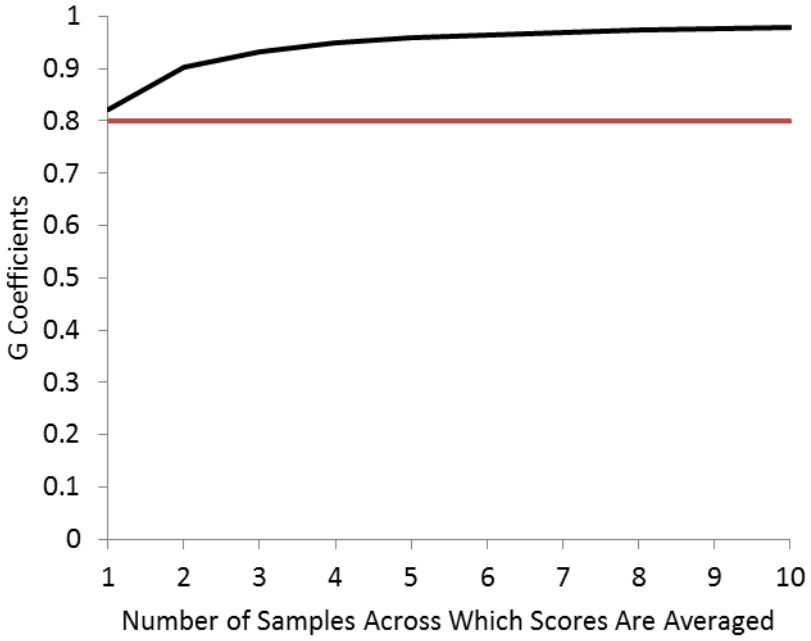


Figure 2. Generalizability coefficients for child vocalization complexity as measured via automated vocal analysis according to collected and projected audio recordings.

However, a single audio recording was not sufficient to obtain a stable estimate of adult word count in the present sample (*g* = .38). Projections suggest that seven day-long audio recordings would be necessary to reach the reliability threshold of .8 that we established a priori (Figure 3).

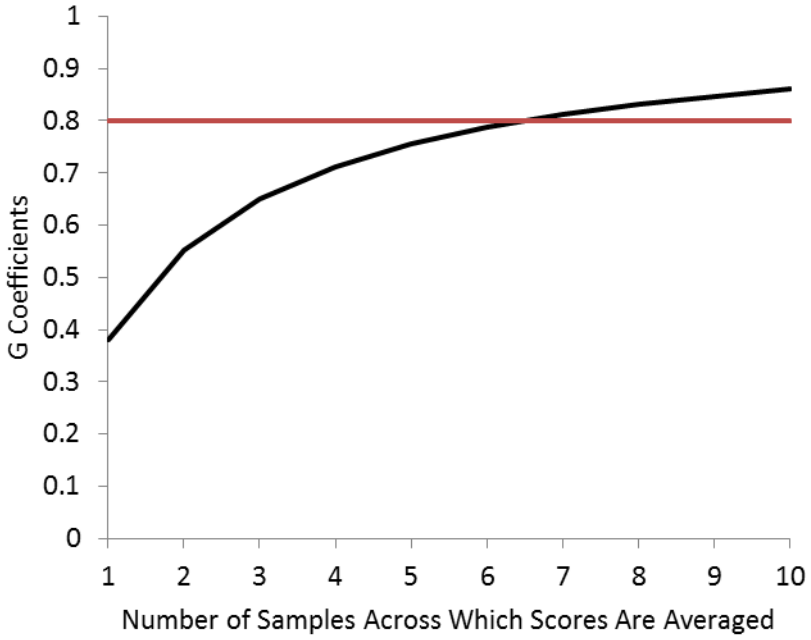


Figure 3. Generalizability coefficients for adult linguistic input as measured via automated vocal analysis according to collected and projected audio recordings.

How Many Conventional Communication Samples are Required to Derive Stable Measures of Child Vocalization Complexity and Adult Linguistic Input?

Our aggregate variable indexing the proportion of communication acts including canonical syllables and the number of different early emerging consonants used communicatively approaches, but does not reach, the .8 threshold for reliability with a single conventional communication sample (*g* = .78). However, an estimate of child vocalization complexity that met our standard for acceptable stability could be obtained by aggregating across two conventional communication samples, as we did in the present study (*g* = .87; Figure 4).

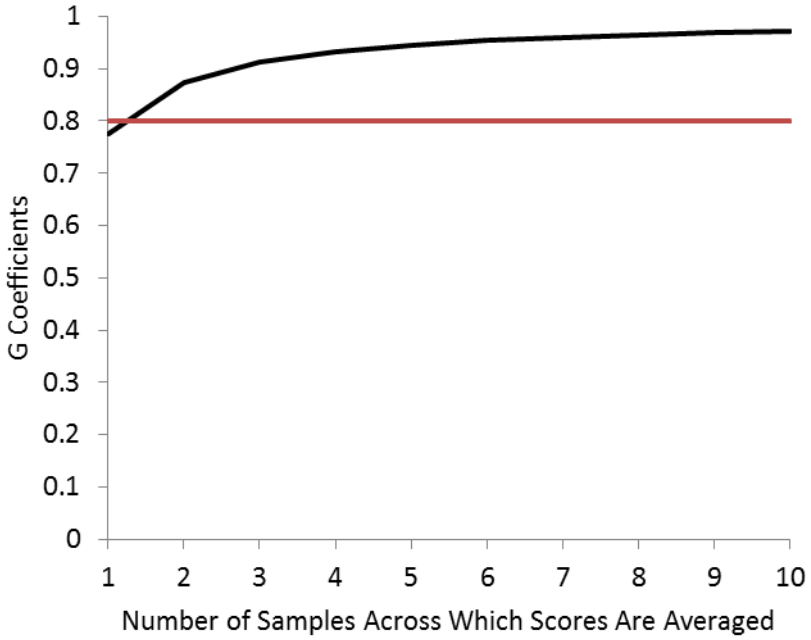


Figure 4. Generalizability coefficients for child vocalization complexity as measured across collected and projected conventional communication samples.

One conventional communication sample was also insufficient to obtain a stable estimate of adult linguistic input in this population (*g* = .56). According to projections, four conventional communication samples would be necessary to achieve a *g* coefficient of .8 for parent linguistic

responses to child attentional and communicative leads in preschoolers with ASD who are still in the early stages of language development (Figure 5). However, it is important to note that some consider a g coefficient of .7 to indicate acceptable stability (Yoder & Symons, 2010). We did attain this less stringent level of stability by aggregating across two parent-child samples in this study (*g* = .71).

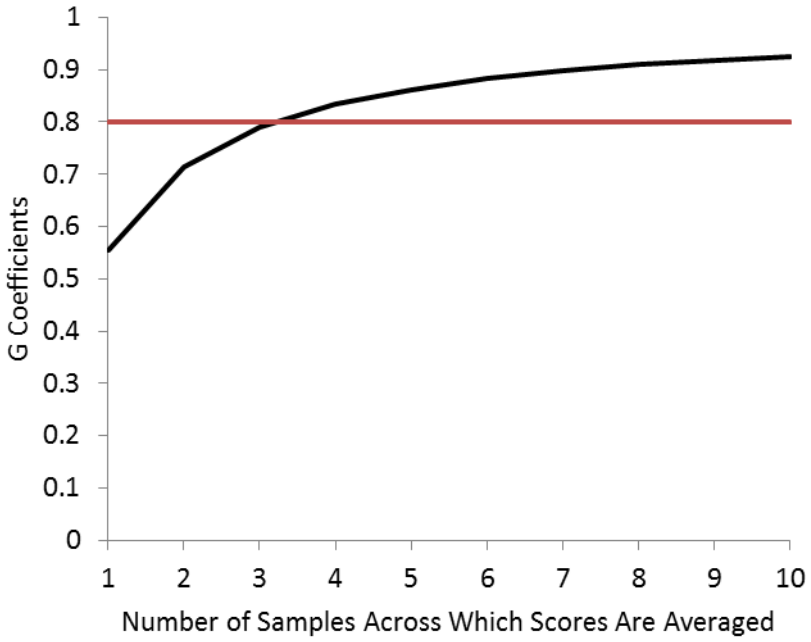


Figure 5. Generalizability coefficients for adult linguistic input as measured across collected and projected conventional communication samples.

Table 5 summarizes the results of stability analyses for indices of child vocalization complexity and adult linguistic input from conventional communication samples and automated vocal analysis.

Table 5

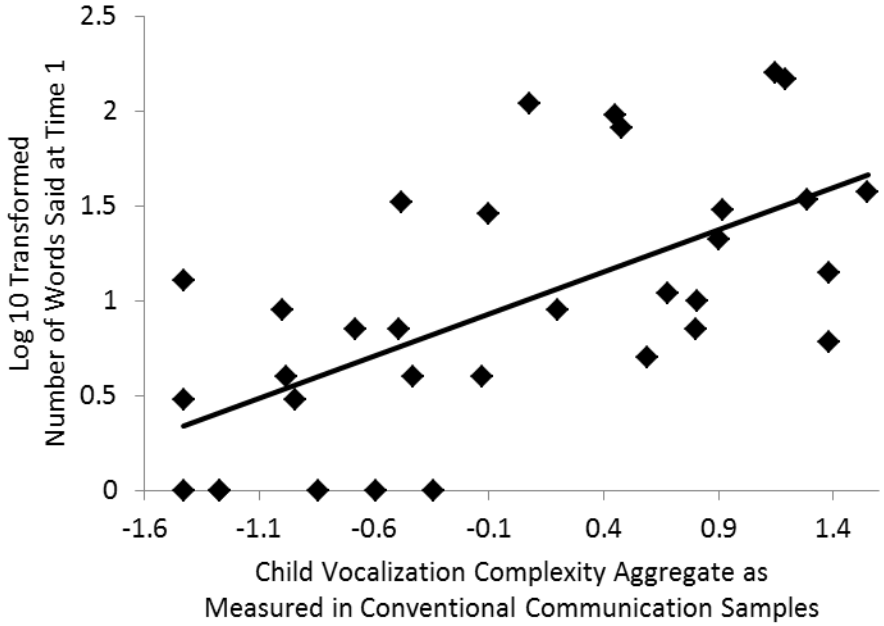
Stability of Indices of Child Vocalization Complexity and Adult Linguistic Input

Variable	*g* observed for two samples	*g* calculated for one sample	# samples for $g \geq .8$

Child Vocalization Complexity from Communication Samples	.87	.78	2
Child Vocalization Complexity from Automated Vocal Analysis	.90	.82	1
Adult Linguistic Input from Communication Samples	.71	.56	4
Adult Linguistic Input from Automated Vocal Analysis	.55	.38	7

Is Child Vocalization Complexity as Measured via Automated Vocal Analysis "Better Than" or at Least "As Good As" Child Vocalization Complexity as Measured in Conventional Communication Samples for Predicting Concurrent and Future Spoken Vocabulary?

Our index of child vocalization complexity as measured in conventional communication samples was significantly correlated with both concurrent ($r(31) = .63$, $p < .001$, one-tailed) and future ($r(31) = .66$, $p < .001$, one-tailed) spoken vocabulary in our sample (Figure 6). The variable indexing child vocalization complexity as derived via automated vocal analysis, the infraphonological vocal complexity score, was also significantly associated with concurrent ($r(31) = .46$, $p = .004$, one-tailed) and future ($r(31) = .51$, $p < .001$, one-tailed) spoken vocabulary (Figure 7). There was not a significant difference between the magnitude of the correlations for our two indices of child vocalization complexity with either concurrent, $Z = -1.119$, $p = .26$, two-tailed, or future spoken vocabulary, $Z = -1.016$, $p = .31$, two-tailed. Thus, we can conclude that our index of child vocalization complexity as derived via automated vocal analysis is at least "as good as" our index of child vocalization complexity as measured in conventional communication samples in explaining variability in concurrent and future spoken vocabulary in preschoolers with ASD who are in the early stages of language development.



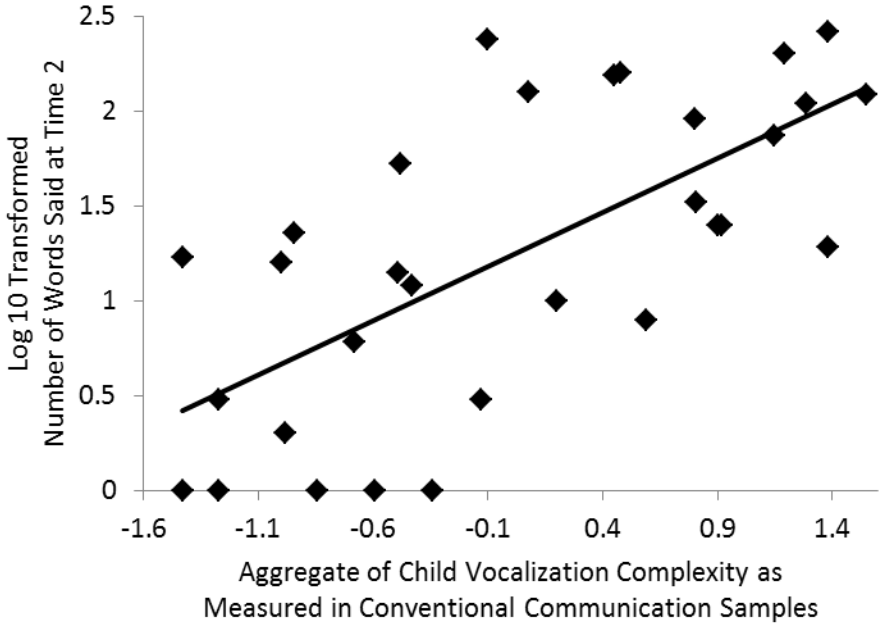


Figure 6. Scatterplots for the associations of child vocalization complexity as measured in conventional communication samples with spoken vocabulary.

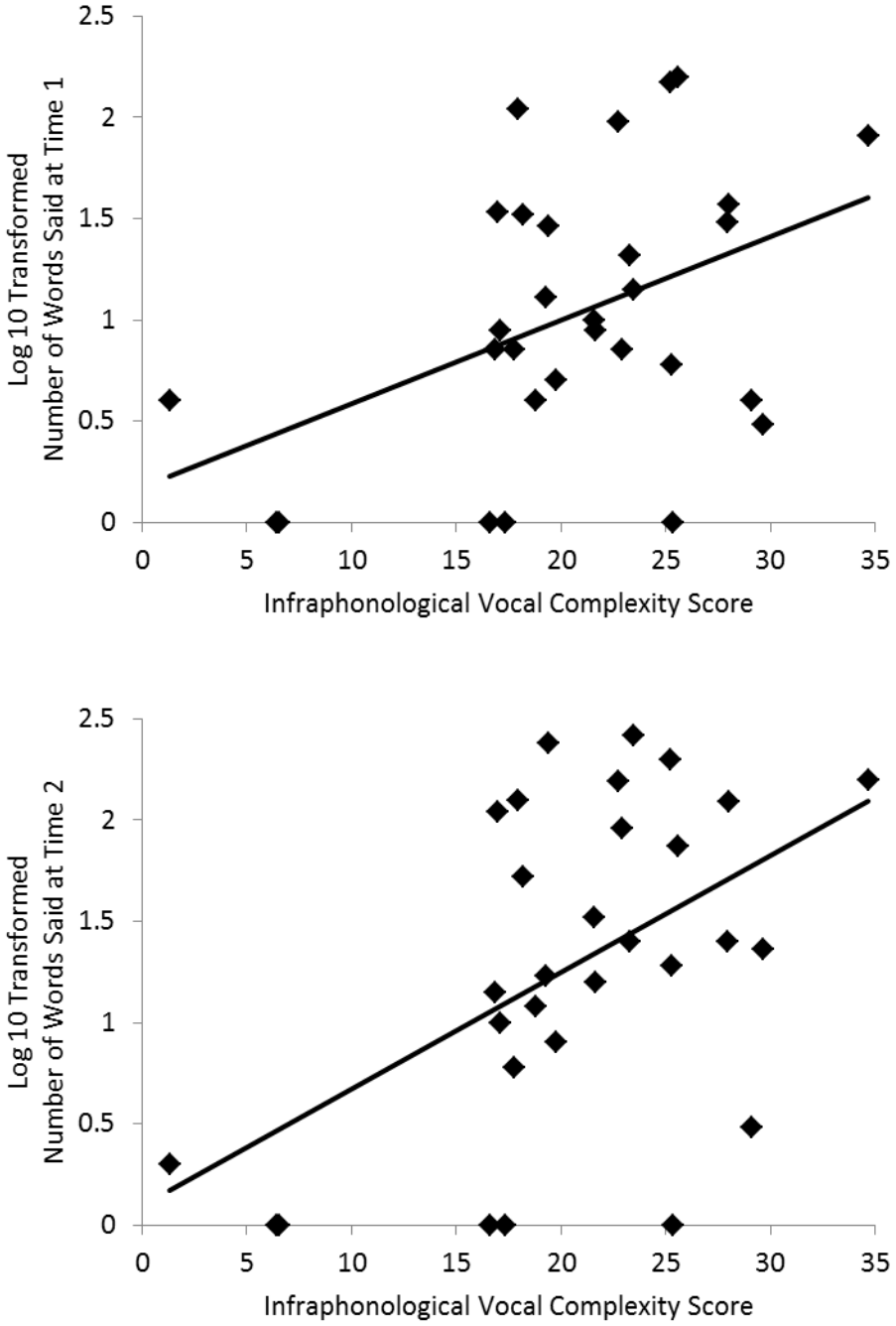


Figure 7. Scatterplots for the associations of child vocalization complexity as derived via automated vocal analysis with spoken vocabulary.

Is Adult Linguistic Input as Derived Via Automated Vocal Analysis "Better Than" or at Least "As Good As" Adult Linguistic Input as Measured in Conventional Communication Samples in Predicting Concurrent and Future Spoken Vocabulary?

We were unable to impute missing values for parent linguistic input as measured in conventional communication samples and as derived via automated vocal analysis because these predictors were not sufficiently associated with any other potential auxiliary variables in our larger data set. Imputing values without sufficient "information" from other variables may unnecessarily increase error variance and reduce statistical power to detect associations of interest (Von Hippel, 2007). Therefore, participants with missing values for adult linguistic input from conventional communication samples and/or automated vocal analysis were excluded list-wise from analyses examining the validity of these variables.

Our index of adult linguistic input as measured in conventional communication samples covaried with concurrent spoken vocabulary ($r(24) = .38$, $p = .03$, one-tailed) and predicted future spoken vocabulary ($r(24) = .38$, $p = .03$, one-tailed) in our sample (Figure 8). However, our proposed index of adult linguistic input as derived via automated vocal analysis, adult word count, was not significantly associated with concurrent ($r(24) = .12$, $p = .27$, one-tailed) or future ($r(24) = .06$, $p = .39$, one-tailed) spoken vocabulary in our sample of preschoolers with ASD (Figure 9). The difference in the magnitude of the correlation coefficients for our two indices of adult linguistic input did not reach statistical significance for either concurrent, $Z = .882$, $p = .38$, two-tailed, or future, $Z = 1.093$, $p = .27$, two-tailed, spoken vocabulary in our sample. This null result was due in part to our relatively small sample size. Thus, there is not evidence that adult linguistic input as measured in conventional communication samples is a significantly better predictor of language than the automated one. There is only evidence that the former has a relation with spoken vocabulary that significantly exceeds zero, while the latter

does not. Thus, the index of adult linguistic input from the conventional communication samples is the only variable that surpasses minimal evidence as a predictor of spoken vocabulary in this population.

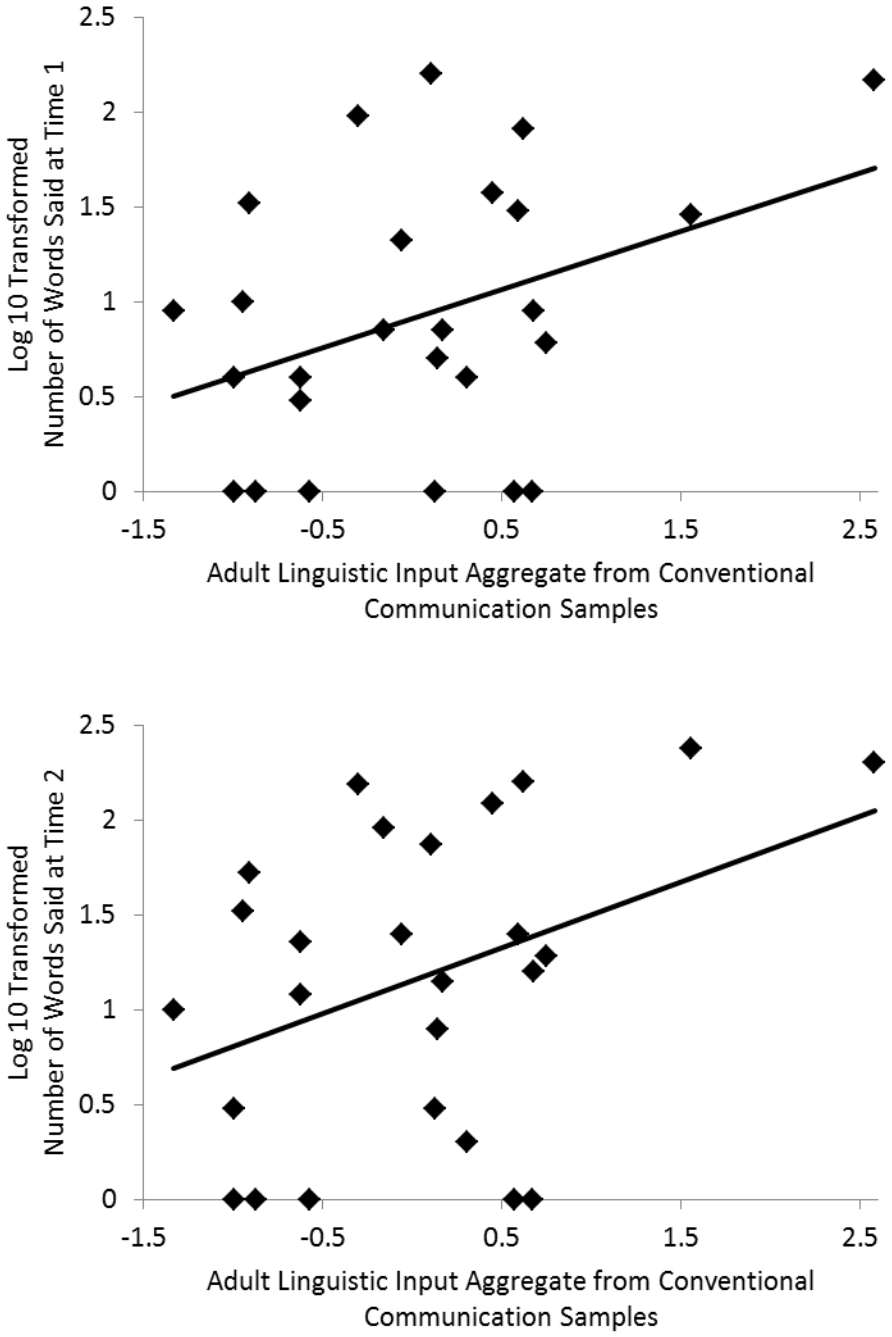


Figure 8. Scatterplots for the associations of adult linguistic input as measured in conventional communication samples with spoken vocabulary.

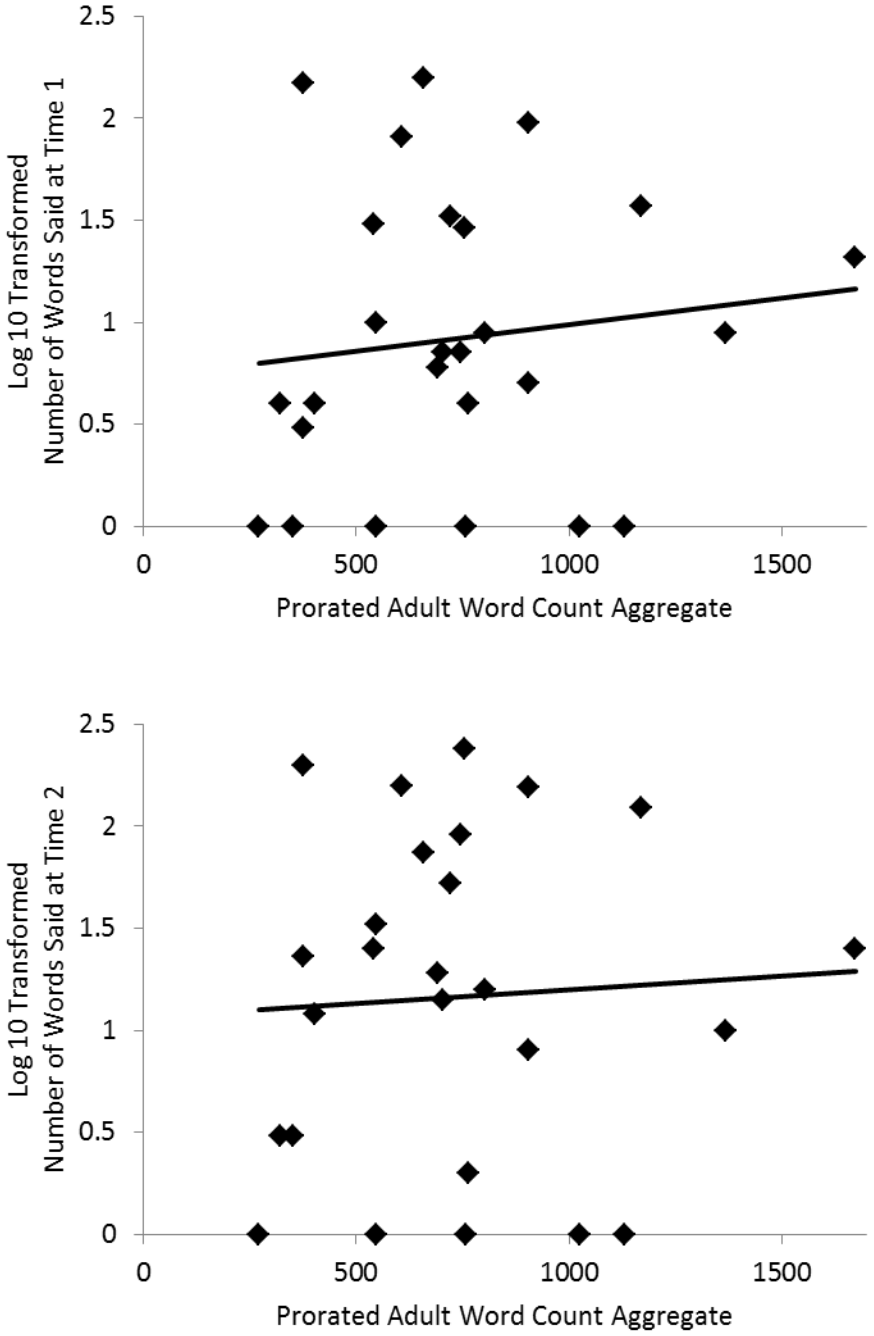


Figure 9. Scatterplots for the associations of adult linguistic input as derived via automated vocal analysis with spoken vocabulary.

Table 6 summarizes the results of our validity analyses for indices of child vocalization complexity and adult linguistic input from conventional communication samples and automated vocal analysis.

Table 6

Validity of Indices of Child Vocalization Complexity and Adult Linguistic Input Variables

Variable	Concurrent Nomological Validity	Predictive Nomological Validity
Child Vocalization Complexity from Communication Samples	.63***	.66***
Child Vocalization Complexity from Automated Vocal Analysis	.46**	.51***
Adult Linguistic Input from Communication Samples	.38*	.38*
Adult Linguistic Input from Automated Vocal Analysis	.12	.06

Note. * $p < .05$. ** $p < .005$. *** $p < .001$.

Post Hoc Analysis: Does Adult Linguistic Input have Added Value in Predicting Concurrent and Future Spoken Vocabulary?

Our primary analyses indicated that our selected index of adult linguistic input as derived via automated vocal analysis was not a valid and reliable alternative to the more established sampling approach to measuring parent linguistic input. This result seemed to suggest that the acquisition and coding of conventional communication samples would be necessary to tap adult linguistic input in a way that is scientifically useful for predicting spoken vocabulary of children with ASD in early stages of

language development. However, it was not clear that adult linguistic input, even when measured by the costly and time-intensive communication sampling method, would improve our predictions of spoken vocabulary once we accounted for children's vocalization complexity. Transactional theory and research involving other typically developing and clinical populations suggest that measures of child vocalization complexity and adult linguistic input are likely to covary. To determine whether acquiring and coding conventional communication samples to derive an estimate of adult linguistic input would have "added value" in predicting spoken vocabulary in our population of interest, we carried out two separate multiple regression analyses: a) one analysis examining the added value of adult linguistic input as measured in conventional communication samples controlling for child vocalization complexity as measured by the same method, and b) one analysis examining the added value of adult linguistic input as measured in conventional communication samples controlling for child vocalization complexity as derived via automated vocal analysis.

Results of the first analysis indicated that child vocalization complexity as measured in conventional communication samples accounted for a significant amount of the variance in concurrent (45%) and future (37%) spoken vocabulary, even when controlling for adult linguistic input as measured in conventional communication samples. However, adult linguistic input as measured in conventional communication samples did not improve predictions of concurrent or future spoken vocabulary after controlling for child vocalization complexity as measured in conventional communication samples (Table 7).

Table 7

Spoken Vocabulary as Predicted by Child Vocalization Complexity and Adult Linguistic Input as Measured in Conventional Communication Samples

Concurrent Spoken Vocabulary						
Predictors	B	S.E.	95% CI	*t*	*p*	part *r*
Constant	.87	.11	[.65, 1.10]	8.15	< .001	
Child Vocalization Complexity from Conventional Communication Samples	.49	.11	[.25, .72]	4.26	< .001	.67
Adult Linguistic Input from Conventional Communication Samples	.14	.12	[-.12, .39]	1.09	.287	.16
Future Spoken Vocabulary						
Predictors	B	S.E.	95% CI	*t*	*p*	part *r*
Constant	1.10	.12	[.85, 1.35]	9.19	< .001	
Child Vocalization Complexity from Conventional Communication Samples	.52	.13	[.26, .79]	4.10	< .001	.61
Adult Linguistic Input from Conventional Communication Samples	.17	.14	[-.11. .46]	1.25	.22	.19

Note. CI = confidence interval.

Results of the second analysis indicated that child vocalization complexity as derived via automated vocal analysis accounted for a significant amount of the variance in concurrent (15%) and future (22%) spoken vocabulary, even when controlling for adult linguistic input as measured in conventional communication samples. However, adult linguistic input as measured in conventional communication samples did not improve predictions of concurrent or future spoken vocabulary after controlling for child vocalization complexity as derived via automated vocal analysis (Table 8). Thus, our

index of adult linguistic input as measured in conventional communication samples did not have added value in predicting concurrent and future spoken vocabulary in preschoolers with ASD in the early stages of language development when we controlled for child vocalization complexity as measured by either conventional communication sampling or automated vocal analysis.

Table 8

Spoken Vocabulary as Predicted by Indices of Child Vocalization Complexity as Derived Via Automated Vocal Analysis and Adult Linguistic Input as Measured in Conventional Communication Samples

Concurrent Spoken Vocabulary						
Predictors	B	S.E.	95% CI	*t*	*p*	part *r*
Constant	.21	.39	[-.60, 1.02]	.54	.59	
Child Vocalization Complexity from Automated Vocal Analysis	.04	.02	[.001,.08]	2.13	.04	.39
Adult Linguistic Input from Conventional Communication Samples	.16	.15	[-.15, .47]	1.09	.29	.20
Future Spoken Vocabulary						
Predictors	B	S.E.	95% CI	*t*	*p*	part *r*
Constant	.15	.42	[-.72, 1.02]	.35	.73	
Child Vocalization Complexity from Conventional Communication Samples	.05	.02	[.01, .09]	2.72	.01	.47
Adult Linguistic Input from Conventional Communication Samples	.18	.16	[-.15, .51]	1.11	.28	.19

Note. CI = confidence interval.

Chapter 4

Discussion

This study drew on extant data from a large-scale study of "useful speech" development in preschoolers with ASD to examine the stability and validity of a novel approach, automated vocal analysis, in the measurement of two previously identified predictors of spoken language in this population – child vocalization complexity and adult linguistic input. Results indicate that this new method may be used to derive an index of child vocalization complexity that is both stable and useful as a predictor of spoken vocabulary for preschoolers with ASD in early stages of language development. In contrast, our findings do not support the stability and validity of our selected index of adult linguistic input as derived via automated vocal analysis. However, adult linguistic input, even as measured in more conventional parent-child interactions, did not improve our predictions of spoken vocabulary once we had accounted for child vocalization complexity. These findings highlight child vocalization complexity as a key predictor of spoken vocabulary and provide support for automated vocal analysis as an alternative to the more costly and time-consuming conventional communication sampling that has historically been used to tap this construct in children with ASD and other developmental disabilities.

Automated Vocal Analysis Yields a Highly Stable Estimate of Child Vocalization Complexity

Our results indicate that automated vocal analysis may be utilized to obtain an estimate of child vocalization complexity that meets high standards for stability. This study extends previous work (Yoder et al., 2013) by demonstrating that a single day-long audio recording is sufficient to obtain a stable infraphonological vocal complexity score in preschoolers with ASD, even when the sample is limited to children on the autism spectrum who are still in the earliest stages of language development. A similarly stable estimate of vocalization complexity may also be obtained with only two conventional communication samples. However, collecting even two conventional communication samples would come at a significantly higher cost in terms of both the time required to acquire and code the samples

and the associated expense relative to acquiring one day-long audio recording, uploading the sample, and automatically generating a score for child vocalization complexity.

In contrast, we did not find that one, or even a few, audio recordings were sufficient to derive a stable estimate for our selected index of adult linguistic input via automated vocal analysis. According to our projections, an entire week's worth of audio recordings would be necessary to obtain a stable estimate of adult word count in our sample. A similarly stable estimate of adult linguistic input may be obtained with relatively fewer conventional communication samples (i.e., four parent-child interactions), but acquiring and coding this many conventional communication samples would come at a very high cost in time and money. In fact, it is probably more feasible for a clinician to collect a week's worth of audio recordings than it is for a clinician to collect and code four conventional communication samples. Linguistic input from communication samples is clearly influenced by number of child communicative or attentional leads. The frequency of such leads is likely to be influenced by toy set, activity, mood, and a variety of other factors. Similarly, adult word count from the automated analysis of day-long samples is likely to be influenced by co-occurring activities and social interactions. Unfortunately, it is difficult to obtain stable estimates for scores that are highly contextually influenced (Yoder & Symons, 2010).

This result has implications for some large-scale projects, such as Providence Talks, Thirty Million Words, and Talk with Me Baby, that are using adult word count as derived via automated vocal analysis in an attempt to measure and change adult linguistic input in various at-risk populations (Hodson, 2014; Robertson & Musso, 2014; Sacks et al., 2013; Suskind et al., 2013). It may be necessary for researchers and clinicians to obtain many more audio recordings than they had originally planned to obtain stable baselines and to capture true change in adult linguistic input. Fortunately, the low cost and effort required to obtain seven consecutive day-long samples with the LENA-DLP makes acquisition of this many samples entirely possible.

Automated Vocal Analysis Is a Nomologically Valid Alternative to Conventional Communication Sampling for Measurement of Child Vocalization Complexity

Our findings further indicate that our selected index of child vocalization complexity as derived via automated vocal analysis provides a valid alternative to conventional communication sampling for the measurement of child vocalization complexity in preschoolers with ASD who are still preverbal or who are just beginning to use words to communicate. The infraphonological vocal complexity score predicted both concurrent and future spoken vocabulary in this population. The present work extends the previous conclusions of Yoder and colleagues (2013), who found that this same index of child vocalization complexity, which they called the vocal development age equivalency score, correlated with concurrent spoken language in a sample of preschoolers with ASD who were largely already using words to communicate.

Our aggregate index of child vocalization complexity as measured in conventional communication sampling contexts also predicted concurrent and future spoken vocabulary in preschoolers with ASD. This is not surprising given that previous studies have found that measures of child vocalization complexity similar to, or even identical to, the component variables comprising our aggregate predict spoken language in preschoolers with ASD who are heterogeneous in terms of spoken language stage (Plumb & Wetherby, 2013; Sheinkopf et al., 2000; Wetherby et al., 2007). Importantly, the infraphonological vocal complexity score was non-significantly different from this aggregate measure of child vocalization complexity as measured across conventional communication samples in its prediction of spoken vocabulary in our sample. It is noteworthy that the index as derived via automated vocal analysis may or may not reflect complexity of vocalizations *used communicatively*. A previous study found that vocalizations used communicatively improved predictions of spoken language in preschoolers with ASD after controlling for vocalizations that were used non-communicatively (Plumb & Wetherby, 2013). However, we found that the effect sizes for the association with concurrent and

future spoken vocabulary were large for both of our indices of child vocalization complexity, regardless of whether or not they captured only vocalizations produced for a communicative purpose.

Using Only Two Day-long Samples, Automated Vocal Analysis Does NOT Appear to Be a Valid Alternative to Conventional Communication Sampling for Measurement of Adult Linguistic Input

Unfortunately, our results do not support the use of automated vocal analysis as an alternative to conventional communication samples for the measurement of adult linguistic input. Our index of adult linguistic input as measured in conventional parent-child samples did predict concurrent and future spoken vocabulary, as expected given previous findings for young children with ASD who are in the earliest stages of language development (e.g., Haebig et al., 2013a, 2013b; McDuffie & Yoder, 2010; Siller & Sigman, 2002, 2008). However, our index of adult linguistic input as derived via automated vocal analysis, adult word count, was simply not correlated with concurrent or future spoken vocabulary in our sample. This was somewhat surprising because recent work by Warren and colleagues (2010) indicated that adult word count as measured with a single audio recording did covary with concurrent spoken language in a sample of children with ASD who were predominantly already using words to communicate. It may be that spoken language level was correlated with adult word count in the Warren et al. study (2010) because adults simply talk more when children speak more. Nonetheless, the present null result suggests that the finding from Warren et al.'s (2010) work does not extend to preschoolers with ASD who are preverbal or just beginning to use words to communicate.

When a priori predictions for theoretically based associations are not confirmed for a variable, we must consider two possible explanations. First, it is possible that the measurement system did not quantify the construct of interest in a reliable or valid manner. Second, it is possible that the theory on which the anticipated associations were based is in need of modification.

Regarding the first possible explanation, the validity of the adult word count variable is certainly limited by the instability that we observed for this score in our sample. We do not have any information

regarding the stability of adult word count in the Warren et al. sample (or in any other sample for that matter), but it is quite possible that the stability of adult talk might be more temporally stable for developmentally older participants, such as those studied by Warren et al. However, it is also possible that this variable simply did not measure this construct in a scientifically useful manner.

The index that did predict spoken vocabulary in our sample specifically tapped adult linguistic input that followed into the child's attentional or communicative lead. This attribute of adult linguistic input might be essential for young children with ASD, especially for those who are just learning to use words to communicate. Children with ASD have difficulty with shifting their attention and following their communication partner's attentional lead, and these difficulties impact their word learning (Baird et al., 2000; Baron-Cohen et al., 1997; Baron-Cohen et al., 1996; Zwaigenbaum et al., 2005).

When adults follow the child's attentional or communicative lead in providing linguistic input, they reduce the demands made on children with ASD. This more readily "processable" linguistic information may be the specific type of linguistic input that supports spoken vocabulary development in preschoolers with ASD who are still in the earliest stages of language development. Adult word count captures all adult linguistic input that is produced near the target child, regardless of whether or not it relates to the child's attentional focus or communicative lead. One recent study that attempted to parse out components of the adult word count score found that only the adult linguistic input that was specifically directed to the child, not adult linguistic input that was simply produced near the child and possibly overheard, predicted future spoken vocabulary in a sample of young children from families of low socioeconomic status (Weisleder & Fernald, 2013). Thus, it may be that the *type* of linguistic input received, rather than simply the *quantity*, is what matters most for many populations.

Accordingly, in regards to the second possible explanation, we do not take these results to suggest that adult linguistic input is not a factor in spoken vocabulary growth of children with ASD who are in the earliest stages of language development. The significant findings for our selected index of

adult linguistic input as measured via conventional communication samples, and for similar indices of adult linguistic input from studies using communication sampling, suggest that this is not the case (e.g., Haebig et al., 2013a, 2013b; McDuffie & Yoder, 2010; Siller & Sigman, 2002, 2008). However, these results add to the evidence suggesting it is not simply the count or "amount" of adult words produced near the child that contributes to word learning. Thus, we may need to adjust our thinking regarding which types of adult linguistic input support spoken vocabulary development for this subgroup of children with ASD, as well as for other children who are at-risk for language delays and disorders. Nonetheless, at present, it does not appear that this particular index of linguistic input as measured via automated vocal analysis is a viable alternative to the indices derived in conventional communication samples.

However, adult linguistic input does not actually account for additional variance in concurrent or future spoken vocabulary in this particular population after controlling for child vocalization complexity. In fact, this is the case regardless of whether we measure children's vocalization complexity using conventional communication sampling or the more novel automated vocal analysis. In other words, once we have obtained a stable and valid measure of children's vocalization complexity, the arduous acquisition and coding of adult linguistic input in these parent-child communication samples would not improve our predictions of spoken vocabulary size for young children with ASD who are still in the early stages of language development. Certainly, this is in part due to the intercorrelation between child vocalization complexity and adult linguistic input. Our indices of child vocalization complexity simply capture some of the variance in spoken vocabulary that is also associated with adult linguistic input. Regardless, this result highlights the role of child vocalization complexity as a key predictor of spoken language in young children with ASD.

Implications for Future Research that May Affect Clinical Practice

This finding lays the foundation for further research into child vocalization complexity in this population. Given that the complexity of child vocalizations has been repeatedly linked to later spoken language in children with ASD (as well as typically developing children and several other clinical populations), we need to determine whether child vocalization complexity is malleable with treatment in young children with ASD who for the most part remain preverbal communicators. Although a number of treatments that target increased vocalization complexity are available (e.g., Prelinguistic Milieu Teaching, Early Start Denver Model, PROMPT), to our knowledge no studies to date have actually measured child vocalization complexity as an intervention outcome in preschoolers with ASD. However, there is evidence that Prelinguistic Milieu Teaching has an effect on proportion of communication acts with canonical syllables in initially nonverbal preschoolers with intellectual disabilities (Woynaroski et al., in press).

Additionally, we may evaluate whether treatment effects on spoken vocabulary or broader spoken language skill are preceded and mediated by earlier effects on vocalization complexity. If a mediation relation is confirmed, child vocalization complexity may provide an early indication of whether a child with ASD is responding to treatment. Again, such indirect relations have been found in another clinical sample of initially nonverbal children (Yoder, Woynaroski, Fey, & Warren, in press).

Finally, child vocalization complexity may moderate response to treatment. For example, we may find that children who enter treatment with below average vocalization complexity have better outcomes or experience greater growth in spoken language when they receive a treatment that specifically targets increased complexity of vocalizations. In contrast, children who already demonstrate above average vocalization complexity may have better outcomes or experience greater growth in spoken language when they receive a treatment that more immediately begins to target word learning and broader spoken language skills. The use of automated vocal analysis to index child vocalization

complexity may significantly reduce the time and cost necessary to conduct this research. However, at present, the software needed to compute the infraphonological vocal complexity score is not part of the standard LENA software package (Yoder et al., 2013).

If made publically available, this more cost-effective and time-efficient method will also make it possible for clinicians to measure child vocalization complexity in everyday clinical practice. The present study, in conjunction with previous research, strongly suggests that child vocalization complexity is useful for forming prognoses about the likelihood, or the extent to which, a child with autism will use words to communicate (Plumb & Wetherby, 2013; Sheinkopf et al., 2000; Wetherby et al., 2007). Clinicians may also utilize automated vocal analysis to evaluate effects of treatment on child vocalization complexity. Additional research, such as the mediation analysis suggested above, is necessary to determine whether increased vocalization complexity actually translates to gains in spoken language. If this mediation relation is confirmed, clinicians may monitor growth in vocalization complexity to index children's progress along the path from prelinguistic to linguistic development.

However, even if increased vocalization complexity is not statistically shown to mediate gains of treatment on later spoken language, more complex vocalizations may still increase the clarity of children's communication acts and thus increase communicative success. Additionally, more speech-like vocalizations may elicit adult linguistic responses, which may further support spoken language development (Woynaroski et al., in press). Accordingly, we consider increased vocalization complexity a worthwhile and functional objective of intervention of its own accord. Thus, this work has several possible implications for the use of automated vocal analysis in both research and clinical practice.

Limitations

This study, however, is not without limitations. First, our longitudinal correlational design does not allow us to infer a causal relation between child vocalization complexity and spoken vocabulary in this population. Although the design did allow us to demonstrate an association between vocalization

complexity and vocabulary size and to demonstrate temporal precedence for the association between early (Time 1) vocalization complexity and later (Time 2) spoken vocabulary, it does not allow us to rule out alternative explanations for this association. Future investigations, such as the treatment studies suggested above that could include important experimental controls and that would permit tests of a mediation relation, will be necessary to confirm that increased vocalization complexity translates to increased spoken vocabulary size. Second, though this study represents a step in the right direction, four months is a relatively short timeframe for prediction of future spoken vocabulary. Subsequent studies should attempt to systematically replicate the present results over longer intervals of time. Third, though the children comprising our sample were developmentally younger than the children participating in previous studies on children with ASD and were still in the earliest stages of language development, some were already using some words to communicate at Time 1. Future work will allow us to ascertain whether the associations observed in the present study extend to other subgroups of children with ASD, such as children who are still preverbal or older children who have remained minimally verbal.

Finally, some may question our decision not to control for Time 1 spoken vocabulary size in predicting Time 2 vocabulary size. We elected not to control for initial vocabulary size because of the large covariation between Time 1 and Time 2 spoken vocabulary. Controlling for Time 1 spoken vocabulary size would have left little variance to explain in our sample and would undoubtedly have rendered our observed predictive association with later language non-significant. However, we do not feel that the strong association between early and later spoken vocabulary invalidates indices of child vocalization complexity as predictors of later spoken vocabulary. Using the unrealistic criteria of attending to associations with language only if significant when Time 1 language is controlled may discourage a potentially important line of research. Furthermore, we are not suggesting that indices of child vocalization complexity are simply additional measures of spoken language. Rather, we are

proposing that these measures of “speech-likeness” of vocalizations provide us with unique insight into a child’s position on the prelinguistic path to spoken vocabulary development and broader spoken language use.

Summary

This research extends previous work by providing evidence for the stability and validity of automated vocal analysis in preschoolers with ASD who are still in the early stages of language development. In this longitudinal correlational study, child vocalization complexity emerged as a key predictor of spoken vocabulary for our sample. Automated vocal analysis was found to yield a stable estimate of vocalization complexity and to provide a valid alternative to conventional communication sampling for measurement of this construct. More work certainly needs to be done, particularly in extending this finding for prediction of future language to other subgroups of children with ASD and predicting vocabulary and broader spoken language use over more extended intervals of time. Nonetheless, this study provides preliminary support for the use of this cost effective and efficient method by researchers interested in increasing our understanding of the role of vocal complexity in language development of children with ASD and by clinicians wishing to make empirically based prognoses regarding which children with ASD are likely to acquire “useful speech”.

References

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders-IV-TR*. Washington, DC: APA.

Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelwright, S., & Drew, A. (2000). A screening instrument for autism at 18 months of age: a 6-year follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(6), 694-702. doi: S0890-8567(09)66238-9 [pii]10.1097/00004583-200006000-00007

Bakeman, R., McArthur, D., Quera, V., & Robinson, B. (1997). Detecting sequential patterns and determining their reliability with fallible observers. *Psychological Methods, 2*(4), 357-370.

Baldwin, D. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*(02), 395-418. doi: doi:10.1017/S0305000900008345

Baron-Cohen, S., Baldwin, D., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development, 68*(1), 48-57. doi: 10.1111/j.1467-8624.1997.tb01924.x

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K., . . . Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. *British Journal of Psychiatry, 168*(2), 158-163.

Bates, E., Dale, P., & Thal, D. (Eds.). (1995). *Individual differences and their implications for theories of language development*. Cambridge, MA: Blackwell.

Billstedt, E., Carina Gillberg, I., & Gillberg, C. (2007). Autism in adults: symptom patterns and early childhood predictors. Use of the DISCO in a community sample followed from childhood. *Journal of Child Psychology and Psychiatry, 48*(11), 1102-1110. doi: 10.1111/j.1469-7610.2007.01774.x

Cohen, J., & Cohen, P. (1984). *Applied Multiple Regression*. Mahwah, New Jersey: Erlbaum.

Collins, L., Schafer, J., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330-351.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt.

Cronbach, L. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Dunham, P., Dunham, F., & Curwin, A. (1993). Joint-attentional states and lexical acquisition at 18 months. *Developmental Psychology, 29*(5), 827-831. doi: 10.1037/0012-1649.29.5.827

Eisenberg, L. (1956). The autistic child in adolescence. *The American Journal of Psychiatry, 112*, 607-612.

Enders, C. (2011a). Analyzing longitudinal data with missing values. [doi:10.1037/a0025579]. *Rehabilitation Psychology, 56*, 267-288. doi: 10.1037/a0025579

Enders, C. (2011b). Analyzing longitudinal data with missing values. *Rehabilitation Psychology, 56*(4), 267-288. doi: 10.1037/a00255792011-22474-001 [pii]

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160. doi: 10.3758/BRM.41.4.114941/4/1149 [pii]

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Fenson, L., Dale, P., Reznick, J., Thal, D., Bates, E., Hartung, J., . . . Reilly, J. (2003). *MacArthur communicative development inventories: User's guide and technical manual*. Baltimore, MD: Paul H. Brookes.

Gillberg, C., & Steffenburg, S. (1987). Outcome and prognostic factors in infantile autism and similar conditions: a population-based study of 46 cases followed through puberty. *Journal of Autism and Developmental Disorders, 17*(2), 273-287.

Goldberg, S. (1977). Social competence in infancy: A model of parent-infant interaction. *Merrill-Palmer Quarterly, 23*, 163-177.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development, 30*(6), 509-516. doi: 10.1177/0165025406071914

Haebig, E., McDuffie, A., & Ellis Weismer, S. (2013a). Brief report: parent verbal responsiveness and language development in toddlers on the autism spectrum. *Journal of Autism and Developmental Disorders, 43*(9), 2218-2227. doi: 10.1007/s10803-013-1763-5

Haebig, E., McDuffie, A., & Ellis Weismer, S. (2013b). The contribution of two categories of parent verbal responsiveness to later language for toddlers and preschoolers on the autism spectrum. *American Journal of Speech Language Pathology, 22*(1), 57-70. doi: 10.1044/1058-0360(2012/11-0004)1058-0360_2012_11-0004 [pii]

Hodson, H. (2014). Automatic voice coach gives conversation tips to parents. *New Scientist, 221*(2954), 22.

Horner, R., Carr, E., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179. doi: 10.1177/001440290507100203

Kobayashi, R., Murata, T., & Yoshinaga, K. (1992). A follow-up study of 201 children with autism in Kyushu and Yamaguchi areas, Japan. *Journal of Autism and Developmental Disorders, 22*(3), 395-411.

Lee, I., & Preacher, K. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer Software]. Retrieved August 1, 2014, from http://quantpsy.org

LENA Research Foundation. (2014). LENA: Every word counts. Retrieved March 16, 2014, from http://www.lenababy.com/LenaHome/why-use-lena-home.aspx?

Locke, J. L. (1996). Why do infants begin to talk? Language as an unintended consequence. *Journal of Child Language, 23*(2), 251-268.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

Lotter, V. (1974). Factors related to outcome in autistic children. *Journal of Autism and Childhood Schizophrenia, 4*(3), 263-277.

McCrae, R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*(1), 28-50. doi: 10.1177/1088868310366253 [pii]

McDuffie, A., & Yoder, P. (2010). Types of parent verbal responsiveness that predict language in young children with autism spectrum disorder. *Journal of Speech Language and Hearing Research, 53*(4), 1026-1039. doi: 10.1044/1092-4388(2009/09-0023)1092-4388_2009_09-0023 [pii]

Mullen, E. M. (1995). *Mullen scales of early learning*: Pearson.

Oller, D. K. (1995). Development of vocalizations in infancy. In H. Winitz (Ed.), *Human communication and its disorders: A review* (Vol. IV, pp. 1-30). Timonium: York Press.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.

Oller, D. K., & Lynch, M. P. (1991). Development of the infrastructure of phonological systems. *The Journal of the Acoustical Society of America, 90*(4), 2295-2295. doi: doi:http://dx.doi.org/10.1121/1.401108

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., . . . Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences, 107*(30), 13354-13359. doi: 10.1073/pnas.1003882107 1003882107 [pii]

Plumb, A., & Wetherby, A. (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)1092-4388_2012_11-0104 [pii]

Reeve, L., Reeve, K. F., Brown, A. K., Brown, J. L., & Poulson, C. L. (1992). Effects of delayed reinforcement on infant vocalization rate. *Jounral of Experimental Analysis of Behavior, 58*(1), 1-8. doi: 10.1901/jeab.1992.58-1

Robertson, M., & Musso, L. (2014). School of nursing collaboration receives United Way Award. *Emory News Center*. Retrieved from http://news.emory.edu/stories/2014/01/nursing_autism_collarboration_award/

Rushton, J., Brainerd, C., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. [doi:10.1037/0033-2909.94.1.18]. *Psychological Bulletin, 94*, 18-38. doi: 10.1037/0033-2909.94.1.18

Rutter, M., Greenfeld, D., & Lockyer, L. (1967). A five to fifteen year follow-up study of infantile psychosis. II. Social and behavioural outcome. *The British Journal of Psychiatry, 113*(504), 1183-1199.

Sacks, C., Shay, S., Repplinger, L., Leffel, K., Sapolich, S., Suskind, E., . . . Suskind, D. (2013). Pilot testing of a parent-directed intervention (Project ASPIRE) for underserved children who are deaf or hard

of hearing. *Child Language Teaching and Therapy, 30*(1), 91-102. doi: 10.1177/0265659013494873

Sandbank, M., & Yoder, P. J. (2014). Measuring representative communication in 3-year-olds with intellectual disabilities. *Topics in Early Childhood Special Education*. doi: 10.1177/0271121414528052

Shavelson, R., & Webb, N. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Sheinkopf, S., Mundy, P., Oller, D. K., & Steffens, M. (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders, 30*(4), 345-354. doi: 10.1023/a:1005531501155

Shumway, S., & Wetherby, A. (2009). Communicative acts of children with autism spectrum disorders in the second year of life. *Journal of Speech Language and Hearing Research, 52*(5), 1139-1156. doi: 10.1044/1092-4388(2009/07-0280)1092-4388_2009_07-0280 [pii]

Siller, M., & Sigman, M. (2002). The behaviors of parents of children with autism predict the subsequent development of their children's communication. *Journal of Autism and Developmental Disorders, 32*(2), 77-89.

Siller, M., & Sigman, M. (2008). Modeling longitudinal change in the language abilities of children with autism: parent behaviors and child characteristics as predictors of change. *Developmental Psychology, 44*(6), 1691-1704. doi: 10.1037/a00137712008-16008-013 [pii]

Snyder-McLean, L. (1990). Communication development in the first two years: A transactional process. *Zero to Three*, 13-20.

Steiger, J. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245-251.

Suskind, D., Leffel, K., Hernandez, M., Sapolich, S., Suskind, E., Kirkham, E., & Meehan, P. (2013). An exploratory study of "quantitative linguistic feedback": Effect of LENA feedback on adult

language production. *Communication Disorders Quarterly, 34*(4), 199-209. doi: 10.1177/1525740112473146

Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.

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, Vol. 1: Diagnosis, development, neurobiology, and behavior (3rd ed.).* (pp. 335-364): Hoboken, NJ, US: John Wiley & Sons Inc.

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., . . . Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech Language and Hearing Research, 52*(3), 643-652. doi: 1092-4388_2009_08-0136 [pii]10.1044/1092-4388(2009/08-0136)

Tapp, J., & Walden, T. (1993). PROCODER: A professional tape control, coding, and analysis system for behavioral research using videotape. *Behavior Research Methods, Instruments, & Computers, 25*(53-56).

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454-1463.

Von Hippel, P. T. (2007). Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology, 37*(1), 83-117. doi: 10.1111/j.1467-9531.2007.00180.x

Warren, S., Gilkerson, J., Richards, J., Oller, D. K., Xu, D., Yapanel, U., & Gray, S. (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders, 40*(5), 555-569. doi: 10.1007/s10803-009-0902-5

Weisleder, A., & Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143-2152. doi: 10.1177/0956797613488145095679761348814 5 [pii]

Wetherby, A., & Prizant, B. M. (2002). *Communication and symbolic behavior scales developmental profile- first normed edition*. Baltimore: Paul H. Brookes.

Wetherby, A., 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. doi: 10.1007/s10803-006-0237-4

Woynaroski, T., Yoder, P., Fey, M., & Warren, S. (in press). A transactional model of spoken vocabulary variation in toddlers with intellectual disabilities. *Journal of Speech, Language, & Hearing Research*.

Xu, D., Yapanel, U., & Gray, S. (2008). Reliability of the LENA language environment analysis system in young children's natural language home environment (LENA Foundation Technical Report LTR-05-2).

Xu, D., Yapanel, U., Gray, S., & Baer, C. (2008). The LENA language environment analysis system: the interpretive time segments (ITS) file (LENA Foundation Technical Report LTR-04-2).

Yoder, P. J., Oller, D. K., Richards, J., 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. doi: 10.1002/aur.1271

Yoder, P. J., & Stone, W. L. (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: 49/4/698 [pii]10.1044/1092-4388(2006/051)

Yoder, P. J., & Symons, F. (2010). *Observational measurement of behavior*. New York, NY: Springer Publishing Company.

Yoder, P. J., Woynaroski, T., Fey, M., & Warren, S. (in press). Why dose frequency affects spoken vocabulary in preschoolers with Down syndrome. *American Journal on Intellectual and Developmental Disabilities*.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience, 23*, 143-152. doi: S0736574804000553 [pii]10.1016/j.ijdevneu.2004.05.001