Research Article

Preliminary Evidence That Growth in Productive Language Differentiates Childhood Stuttering Persistence and Recovery

Kathryn A. Leech,^{a,b} Nan Bernstein Ratner,^a Barbara Brown,^c and Christine M. Weber^c

Purpose: Childhood stuttering is common but is often outgrown. Children whose stuttering persists experience significant life impacts, calling for a better understanding of what factors may underlie eventual recovery. In previous research, language ability has been shown to differentiate children who stutter (CWS) from children who do not stutter, yet there is an active debate in the field regarding what, if any, language measures may mark eventual recovery versus persistence. In this study, we examined whether growth in productive language performance may better predict the probability of recovery compared to static profiles taken from a single time point.

Method: Productive syntax and vocabulary diversity growth rates were calculated for 50 CWS using random coefficient models. Logistic regression models were then used to determine

whether growth rates uniquely predict likelihood of recovery, as well as if these rates were predictive over and above currently identified correlates of stuttering onset and recovery.

Results: Different linguistic profiles emerged between children who went on to recover versus those who persisted. Children who had steeper productive syntactic growth, but not vocabulary diversity growth, were more likely to recover by study end. Moreover, this effect held after controlling for initial language ability at study onset as well as demographic covariates.

Conclusions: Results are discussed in terms of how growth estimates can be incorporated in recommendations for fostering productive language skills among CWS. The need for additional research on language in early stuttering and recovery is suggested.

Developmental stuttering is a communication disorder affecting 4%–8% of children at some point during early childhood (Bloodstein & Bernstein Ratner, 2008; Yairi & Ambrose, 2013). It is well known that the majority of children who begin to stutter eventually outgrow this condition, with spontaneous recovery rates estimated at levels of 80% or higher (Reilly et al., 2013; Yairi & Ambrose, 1999). However, the majority of children who do not recover face significant life impacts. Stuttering that becomes a chronic lifelong condition is associated with a host of adverse consequences, from diminished psychosocial adjustment and well-being (McAllister, Collier, & Shepstone, 2013) to diminished vocational and economic potential (Rees & Sabia, 2014). These impacts are lifelong,

from early childhood to old age (Bricker-Katz, Lincoln, & McCabe, 2009).

Factors Predicting Likelihood of Stuttering Recovery

It is of both practical and theoretical interest to identify factors that may be associated with increased odds of spontaneous recovery from stuttering. A number of factors have been implicated in estimating profiles likely to characterize children who recover from stuttering. These range widely among familial, behavioral, and neurophysiological indices. Risk of persistence is conveyed by family history of persistent, rather than recovered, stuttering (Ambrose, Cox, & Yairi, 1997); differences in measures of cortical grey matter in the superior temporal gyrus and white matter integrity along the arcuate fasciculus (Chang, Erickson, Ambrose, Hasegawa-Johnson, & Ludlow, 2008); and motor abilities (Ambrose, Yairi, Loucks, Seery, & Throneburg, 2015; Walsh, Mettel, & Smith, 2015).

Demographic markers have also been implicated in differentiating children who are at risk for persistent stuttering versus those who are more likely to recover. Gender

^aUniversity of Maryland, College Park

^bHarvard Graduate School of Education, Cambridge, MA

^cPurdue University, West Lafayette, IN

Correspondence to Kathryn A. Leech: kleech@bu.edu

Editor: Julie Liss

Associate Editor: Hans-Georg Bosshardt

Received September 19, 2016

Revision received March 6, 2017

Accepted May 27, 2017

https://doi.org/10.1044/2017_JSLHR-S-16-0371

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

appears to mark both the likelihood of receiving a stuttering diagnosis (Bloodstein, 1995; Yairi & Ambrose, 1992a) as well as persistent stuttering (Dworzynski, Remington, Rijdsdijk, Howell, & Plomin, 2007; Yairi & Ambrose, 1992b). Furthermore, there is some indication that children who receive a stuttering diagnosis at later ages are less likely to recover (Yairi, Ambrose, Paden, & Throneburg, 1996; see discussion in Yairi & Ambrose, 2005), though this finding is confounded by the fact that stuttering onset for boys occurs later than for girls, on average, and may reflect differential rates of language maturation by gender (Yairi & Ambrose, 1992a). Parental education and socioeconomic status have also been identified as potential factors in elevating risk for stuttering onset in a small set of studies with contradictory findings (e.g., Boyle et al., 2011; Richels, Johnson, Walden, & Conture, 2013; Watts, Eadie, Block, Mensah, & Reilly, 2015), though it is currently unknown whether socioeconomic status may differentiate stuttering recovery versus persistence.

The Possible Role of Language in Stuttering Onset and Recovery Likelihood

An ever-growing literature has identified a variety of moderately, subclinically depressed language skills in the general population of children who stutter (CWS; see extensive summaries in Bloodstein & Bernstein Ratner, 2008; Hall, Wagovich, & Bernstein Ratner, 2007). However, the generalizability and extent of this finding have been seriously questioned, most notably in reviews of the literature by Nippold (1990, 2012). Three studies from the same research group and child cohort have found the inverse, that very young CWS have equivalent or better skills than children who do not stutter (CWNS; Reilly et al., 2013; Watts, Eadie, Block, Mensah, & Reilly, 2017; Watts et al., 2015). There are also studies finding no difference between the language abilities of CWS and CWNS, complicating the picture even further (e.g., Bernstein Ratner & Sih, 1987; Nippold, Schwarz, & Jescheniak, 1991).

Given that a proportion of studies involving CWS had relatively small sample sizes, meta-analyses may reveal trends not as readily discerned in smaller, separate reports. A meta-analysis conducted by Ntourou, Conture, and Lipsey (2011) concluded that CWS scored significantly lower than CWNS on a broad array of indices of both vocabulary and syntactic maturity, including standardized batteries measuring overall language skill, on measures of receptive language and specifically receptive vocabulary, on measures of expressive vocabulary, and mean length of utterance (MLU) in spoken language samples. The largest general depression across studies was seen for measures of receptive grammar, with the widest variation in reported scores seen for expressive grammar. Critically and perhaps contributing to the controversy, CWS do not generally score below the typically developing range on these assessments, suggesting very subtle delays in language not sufficient to trigger a need for intervention. Thus, the implications

of this potential relative developmental disadvantage are not yet clear.

More germane to the current investigation, beyond general reports of language ability in cohorts of CWS, a large proportion of studies have examined specific linguistic predictors of *persistence* and *recovery*. Because 70%–80% of children are likely to experience spontaneous recovery from stuttering, it is of both practical and scientific value to determine whether or not there are features of early stuttering that can inform recovery. These studies have used a variety of metrics, including language proficiency scores close to onset (Yairi et al., 1996) and phonological encoding skill, including nonword repetition ability (Spencer & Weber-Fox, 2014) and phonological skills measured by standardized assessments (Paden & Yairi, 1996; Paden, Yairi, & Ambrose, 1999), dissociations among scores on standardized test measures (Clark, Conture, Walden, & Lambert, 2015), expressive language measures (Watkins & Yairi, 1997; Watkins, Yairi, & Ambrose, 1999), standardized language assessment scores over time (Ambrose et al., 2015), cortical processing indices of speech and language (such as event-related potentials; Mohan & Weber, 2015; Usler & Weber-Fox, 2015; Weber-Fox, Wray, & Arnold, 2013), and temperamental factors (Ambrose et al., 2015).

Language Growth as a Predictor of Recovery Status

To date, two of the largest and longest prospective longitudinal studies of childhood stuttering have been conducted by the University of Illinois Stuttering Project and the Purdue Stuttering Project. These studies differ in some critical ways from the sample of children identified and followed as part of a larger Australian epidemiological study of general child health and development (Reilly et al., 2013; Watts et al., 2015). It is notable that the majority of reports from both of the American sites concur on basic risk factors for persistence, in a combination of motor and linguistic profiles, particularly when combined with genetic risk. These studies identified potential differences between CWS and typically fluent peers, as well as markers for persistence rather than recovery during childhood.

Although promising, the majority of these studies examined static profiles (e.g., intake measures at first visit) as predictors of eventual fluency status, rather than considering how more dynamic measures (e.g., change over time) in linguistic profiles may also be associated with the probability of recovery. Thus, differences between groups, particularly between children whose stuttering persists versus children whose stuttering recovers, may be more clearly delineated in growth or change over time rather than in performance at one specific time point. In this regard, statistical procedures such as growth modeling can provide additional clues about recovery profiles by isolating initial skill sets that mark risk for persistence or recovery and mapping trajectories toward eventual fluency status. The need to examine growth, rather than skills at onset, is underscored by the fact that children's lexical and syntactic skills are rapidly developing during the time that stuttering has its

typical onset, between 2;6 and 3;6 years;months (Bloodstein, 2006). Historically, numerous studies have noted that stuttering is exacerbated during the production of syntactically demanding utterances (e.g., Bauerly & Gottwald, 2009; Bernstein Ratner, 1998; Gaines, Runyan, & Meyers, 1991; Watson, Byrd, & Carlo, 2011). Kinematic studies of the effects of syntactic complexity on fluent speech movement variability have reported that young CWS often cannot even comply with experimental tasks to fluently produce enough trials of longer, more complex utterances (MacPherson & Smith, 2013). These findings raise the possibility that initial stuttering risk may be elevated by concomitant language abilities at the time that stuttering emerges and that *growth* in productive language proficiency may somehow mark the likelihood of recovery or persistence.

The Current Study

Language abilities have long been implicated in stuttering during the childhood period, yet the exact nature of this relationship is not fully understood, and there is active debate regarding the relative language skills of CWS, as well as those of children who do and do not recover normal fluency. Instead of examining group differences in language performance at individual time points as has been done in past research, this study examined whether *growth* over time reveals important distinctions between stuttering recovery and persistence. Specifically, the main goal of the study was to determine whether language production growth rates over a 3-year period predict the likelihood of stuttering recovery among a sample of preschool-aged children who held a clinical stuttering diagnosis at the beginning of the 3-year period. To measure growth, we employed a statistical approach to estimate the rate of productive language growth using random coefficient models. Studies with typically fluent children have shown growth estimates of productive vocabulary using this method to significantly predict later vocabulary skill (Rowe, Raudenbush, & Goldin-Meadow, 2012). To our knowledge, however, this analytic strategy has yet to be used to predict later outcomes among children with clinical diagnoses related to fluency.

While expanding the definition of language production from static measures to dynamic measures of growth, we narrowed our focus to two features of language production: vocabulary diversity and syntactic complexity. Given that we are among the first to focus on productive language growth among clinical samples of children, we were guided by previous research in determining which measures of vocabulary and syntax to choose as a starting point for the analysis. We followed a recent report by Hollister, Van Horne, and Zebrowski (2016) by measuring syntactic complexity using the Index of Productive Syntax (IPSyn). Among children who eventually recovered from stuttering, Hollister and colleagues found that higher IPSyn scores were associated with reduced likelihood of producing a disfluency. However, this association was not observed for children whose stuttering persisted. Though Hollister and colleagues did not find significant group differences in

IPSyn scores at individual time points, our hypothesis is that *growth* in IPSyn may be more effective at distinguishing recovery status. We also explored the predictive capacity of vocabulary, as work with typically fluent children has found that pace of vocabulary development during the early childhood period is predictive of later outcomes such as language ability at kindergarten entry (Rowe et al., 2012). We chose VOCD as our measure of vocabulary diversity (Malvern & Richards, 1997) because it is more sensitive to age-related change among children in our sample age range (Lai & Schwanenflugel, 2016) and is less impacted by differences in the size of the speech sample (Malvern, Richards, Chipere, & Durán, 2004; Owen & Leonard, 2002).

If measures of language growth do indeed predict recovery status, a second study goal was to determine if this effect holds after controlling for previously documented correlates of stuttering recovery. Specifically, we included demographic factors previously shown to be related to stuttering diagnosis and standardized measures of verbal and nonverbal language in a prediction model with language growth rates. To our knowledge, these variables have never been explored together in a single analysis of CWS, and this approach allows us to determine the unique effect of each factor on recovery likelihood. Our specific research questions were as follows:

1. Does growth in syntactic diversity, as measured by IPSyn, predict the likelihood of recovery from stuttering during the preschool period?
2. Does growth in vocabulary diversity, as measured by VOCD, predict the likelihood of recovery from stuttering during the preschool period?
3. If so, do these associations hold once controlling for currently identified correlates of stuttering onset and recovery?

Method

Participants

Seventy-four children (52 boys, 22 girls) participated in the study, which was conducted at Purdue University and University of Iowa as part of the Purdue Stuttering Project. At the start of the study, 24 children were classified as CWNS, and the remaining 50 were diagnosed as CWS. Table 1 provides demographic information at the group level. The average age of children at the first visit to the laboratory was 57 months, 71 months at Visit 2, and 83 months at Visit 3. Of the 74 participants, 60 participated at all three visits, three missed one visit, and 11 missed two visits. One of the advantages of random coefficient models is that they can handle missing data, allowing us to include all participants in the present analysis (Singer & Willett, 2003).

Group Classification

All CWS were stuttering at the first visit as evaluated by a speech-language pathologist. To be classified as CWS, a child met the following criteria established by Yairi and

Table 1. Means (standard deviations) of demographic variables and standard assessment performance by group.

	CWS-P	CWS-R	<i>p</i> Value
Demographic factors			
Age at enrollment (in months)	62	57	.11
Stuttering onset (in months)	38.20 (10.90)	35.96 (11.72)	.63
Maternal education	5.60 (1.10)	6.11 (0.92)	.08*
Number of boys	15	11	.43
Standardized assessments			
TACL-3	111.45 (15.35)	114.64 (12.94)	.43
SPELT-3	98.32 (12.97)	100.54 (14.92)	.58
BBTOP	82.27 (13.21)	96.68 (12.76)	.001***
Nonword repetition	68.58 (11.87)	77.15 (10.48)	.01**

Note. CWS = children who stutter; P = persistent; R = recovered; TACL-3 = Test of Auditory Comprehension of Language—Third Edition; SPELT-3 = Structured Photographic Expressive Language Test—Third Edition; BBTOP = Bankson-Bernthal Test of Phonology.

* $p < .1$. ** $p < .01$. *** $p < .001$.

Ambrose (1999): (a) parent report of a stuttering problem, (b) speech-language pathologist report of a stuttering problem, (c) rated by either a parent or speech-language pathologist as having a stuttering severity of 2 or higher on an 8-point scale (0 = *normal* to 7 = *very severe*), and (d) exhibited at least three stuttering-like disfluencies per 100 syllables of spontaneous speech obtained during two language samples.

Procedure

Children visited the laboratory once a year for 3 years. At the first visit, children were administered a standardized assessment battery, described in more detail below. Spontaneous speech samples elicited using a standardized protocol were also taken at each visit during free-play with either the parent or a clinician. Each sample was transcribed and coded twice, first by a trained graduate student in speech-language pathology and then by the third author to assure reliable parsing of utterance boundaries. The unit of transcription was the level of the utterance, defined as any verbal utterance bounded by grammatical closure, terminal intonation contour, or a pause of more than 2 s by the speaker. If there were discrepancies in coding, the transcribers reviewed the sample until agreement was reached by consensus. This portion of analysis was conducted at Purdue University. Transcripts were then audited a third time, at the University of Maryland to ensure that any misspellings that might impact vocabulary diversity measures were rectified. Samples were initially transcribed using Systematic Analysis of Language Transcripts (Miller & Chapman, 1984) and converted to Codes for the Human Analysis of Transcripts/Computerized Language Analysis (CLAN) conventions of the Child Language Data Exchange System (MacWhinney, 2000), so that analyses to yield the dependent measures of interest could be performed. The final corpus of language from the children in the study contained 53,065 utterances and 239,698 words.

Measures

Dependent Variable: Clinical Judgment of Stuttering Recovery

The outcome of interest was a dichotomous variable indicating whether the child's stuttering recovered or persisted. The clinical judgment of stuttering recovery was dummy coded (0 = *persisted*, 1 = *recovered*). At each of the three visits, children's fluency was reassessed to determine possible recovery status. A child who stuttered was classified as recovered if the following criteria were met: (a) the speech-language pathologist judged that the child did not exhibit stuttering, (b) the speech-language pathologist rated stuttering severity as less than 2 (on an 8-point scale, on which lower scores were considered less severe), (c) the parent judged that the child did not exhibit stuttering, (d) the parent rated stuttering severity as less than 2, and (e) stuttering-like disfluencies occurred in fewer than three per 100 syllables in the spontaneous language samples. Of the 50 children who were initially classified as CWS, 28 children demonstrated persistent stuttering (persistent; CWS-P), and 22 met criteria for recovery at either the second or third visit (recovered; CWS-R). Of note, no significant differences were observed between groups in age of study enrollment, $t(42) = 1.72$, $p = .11$. Furthermore, no children were reported to relapse back from recovery to stuttering status across the three visits.

Demographic Covariates

Three demographic factors were included in the analyses that have been demonstrated in the extant literature to be associated with stuttering and the likelihood of stuttering recovery: maternal education, age of stuttering onset, and gender. Descriptive statistics for each variable are presented in Table 1.

Maternal education. Maternal education was measured by the number of years of formal schooling. At study onset, mothers self-reported their highest level of education using the Hollingshead Index (Hollingshead, 1975/2011),

a 7-point scale (1 = *less than seventh grade* to 7 = *graduate or professional training*) of various levels of educational attainment. Mothers of children in the sample were highly educated, receiving at least a 4-year college degree on average ($M = 6.02$, $SD = 1.02$, range: 3.0–7.0). This variable served as a proxy for children's socioeconomic status. A one-way analysis of variance (ANOVA) revealed a marginally significant difference in maternal education by group at study onset, $t(48) = 1.81$, $p = .08$. The nature of the difference was such that CWS-P ($M = 5.60$, $SD = 1.10$) had mothers with fewer years of education than CWS-R ($M = 6.11$, $SD = 0.92$).

Age of stuttering onset. For CWS, parents reported the age (in months) in which their child's stuttering began. On average, children's stuttering appeared at 37.2 months ($SD = 11.18$, range: 20–80 months). Parental report of stuttering onset did not significantly differ between CWS-R ($M = 35.96$, $SD = 11.72$) and CWS-P groups ($M = 38.20$, $SD = 10.90$), $t(46) = 0.89$, $p = .63$.

Gender. Gender was similarly distributed between CWS conditions, $\chi^2(1) = 0.62$, $p = .43$. Specifically, there were 15 boys in the CWS-P group and 18 boys in the CWS-R group. Thus, at least when assessed during this age range (nearly 5 years old at Time 1), gender did not appear to predict likelihood recovery, as has been reported at younger ages.

Standardized Language Assessments

At study onset, experimenters assessed children's verbal and nonverbal abilities on a battery of standardized measures. Descriptive statistics by group can be found in Table 1. The Test of Auditory Comprehension of Language–Third Edition (TACL-3; Carrow-Woolfolk, 1999) measured receptive language, and the Structured Photographic Expressive Language Test–Third Edition (SPELT-3; Dawson, Stout, & Eyer, 2003) assessed expressive language abilities. The Bankson-Bernthal Test of Phonology (BBTOP; Bankson & Bernthal, 1999) measured phonological and articulation abilities, and standardized scores from this measure were used in the analyses. Finally, the Dollaghan and Campbell Nonword Repetition Test (Dollaghan & Campbell, 1998) was used to assess auditory perception, phonological working memory, and speech motor planning and execution by asking children to repeat unfamiliar nonwords of increasing length (one to four syllables). The average percentage of correct phonemes across one-, two-, three-, and four-syllable nonwords was used as the measure of nonword repetition in all analyses. Data from five children on the nonword repetition test were not available ($n = 3$ in recovered group, $n = 2$ in persistent group), and thus, any analyses including this measure were run on a sample of 45 children. Data from all children were available for each of the other standardized measures.

Measures of Productive Syntax and Vocabulary Diversity

This study measured productive syntax using IPSyn (Scarborough, 1990) and vocabulary diversity using VOCD (Durán, Malvern, Richards, & Chipere, 2004). The Kideval

program in CLAN automatically calculates these values, making interrater reliability calculations unnecessary. Table 2 displays descriptive statistics for both measures at each of the three time points.

IPSyn. IPSyn assesses the mastery of 56 different morphological and syntactic structures using a 100-utterance sample from each transcript. IPSyn was chosen as the measure of syntactic complexity because it provides a more nuanced assessment of expressive language maturity than does a simple measure of length, such as MLU. Brown (1973), who created the measure, cautioned that MLU loses much of its informative value past the point where average MLU exceeds four morphemes. MLU in our sample exceeded Brown's cautionary threshold; on average, children's MLU at study onset was 4.87 ($SD = 1.12$) and grew approximately 0.5 morphemes with each subsequent visit. IPSyn, in contrast to MLU, has the ability to detect even fine differences in expressive language achievement between children with and without a history of late talking (Rescorla & Turner, 2015).

VOCD. Lexical diversity in spoken language is taken to be an index of the child's available expressive vocabulary in a given context, rather than response to a standard set of predetermined prompts. We chose VOCD (Durán et al., 2004) as the measure of vocabulary diversity over other measures such as number of different words or type-token ratio (TTR; e.g., Cook, Donlan, & Howell, 2013). Although some previous studies of CWS have used TTR to measure vocabulary diversity (e.g., Cook et al., 2013), this measure is known to be impacted by sample size (Owen & Leonard, 2002), which varied between children in our sample. Recent investigations suggest that context shift (which was relatively stable during language sampling for this cohort) is the major likely contaminant for VOCD (Fergadiotis, Wright, & West, 2013). In addition, VOCD has been shown to correlate significantly with standardized vocabulary test scores in CWS (Silverman & Bernstein Ratner, 2002). It also shows a clear pattern of growth over childhood, unlike TTR (Bernstein Ratner & MacWhinney, 2016).

Calculating productive language growth rates. To test our hypothesis that growth in language skills provides information about the CWS's likelihood of recovery, we computed the pace (in months) at which each child's IPSyn and VOCD grew over the 3-year study period. We then used these growth rates as independent variables to predict the likelihood of stuttering recovery. Data from all children in the sample, both CWS and CWNS, were used to calculate growth rates. However, we did not include CWNS in these regression models because we were interested in what factors predict stuttering recovery, which assumes that the child had a clinical diagnosis in the first place.

Growth rates, sometimes referred to as empirical Bayes' posterior means (Raudenbush & Bryk, 2002, Chapter 6), were calculated using the random effects and fixed effects coefficients from a random slope coefficient model (Raudenbush & Bryk, 2002). An estimate was created for each child, computed from a weighted combination of the individual child's own growth trajectory (the random effect

Table 2. Descriptive statistics (mean, standard deviation, and range) for observed VOCD and IPSyn scores at each time point.

Group	VOCD			IPSyn			Utterances		
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
CWS-R									
Visit 1	78.05	17.81	27–109	81.00	8.28	65–97	298.89	96.30	118–571
Visit 2	85.99	11.13	53–109	86.26	6.75	68–98	282.37	94.30	130–581
Visit 3	87.68	17.97	49–123	88.63	5.39	79–98	258.28	100.14	75–564
CWS-P									
Visit 1	81.88	19.68	32–109	81.47	6.33	71–96	311.61	114.01	113–523
Visit 2	83.74	13.04	66–109	82.82	8.13	67–97	263.76	96.80	137–430
Visit 3	84.54	15.58	59–117	85.00	8.08	63–96	224.81	71.63	102–334

Note. VOCD = measure of vocabulary diversity (Malvern & Richards, 1997); IPSyn = Index of Productive Syntax; CWS = children who stutter; P = persistent; R = recovered.

coefficient) as well as the average trajectory of the entire sample (the fixed effect coefficient). Rowe and colleagues (2012) successfully used this approach to show that growth in expressive vocabulary (ages 14–46 months) among a typically fluent sample predicted a standardized measure of receptive vocabulary at kindergarten entry. We chose this approach to estimating growth rates because these estimates are often more precise than differences scores in growth or raw estimates of language production from single time points (Raudenbush & Bryk, 2002, Chapter 6).

The growth rates were obtained by building a model described in Equations (1) and (2). We built two of these models, one to describe growth in VOCD and the other to describe IPSyn growth; the model for VOCD is defined in Equations (1) and (2) and was identical for IPSyn.

$$\text{VOCD}_{it} = \pi_{0i} + \pi_{1i}(t_{it} - 57) + e_{it} \quad e_{it} \sim N(0, \sigma^2) \quad (1)$$

$$\pi_{0i} = \gamma_{00} \quad (2)$$

$$\pi_{1i} = \gamma_{10} + u_{1i} \quad u_{1i} \sim N(0, T)$$

In Equation (1), we model change in VOCD (and IPSyn) as linear in time, given that descriptive statistics showed that the majority of children in the sample demonstrated positive linear growth over the 3 years of study observation (see Table 2 for descriptive statistics by time point). We chose to represent time (*t*) as children's age (in months) because children enrolled in the study at varying ages and other metrics (e.g., per year) may be too coarse-grained to detect change over time. Given this decision, we centered age around the average age at enrollment (mean age = 57 months), allowing parameters to be more interpretable according to Singer and Willett (2003). If we did not center age, the model would estimate growth rates when children are aged 0 months, for which we would expect no measureable productive language skill or growth. This centering choice therefore estimated children's VOCD at 57 months and rate of change at 57 months, regardless of when children actually entered the study. Admittedly, we would have estimated growth closer to the onset of stuttering; however, the study design did not include spontaneous speech data that proximal to onset (average age = 37 months).

This limitation will be further discussed in the Discussion section below.

Equation (1) says that VOCD for child *i* at time *t* can be predicted by the age of child *i* at time *t* (*t_{it}*), VOCD score at 57 months (π_{0i}), and the rate of linear change in VOCD at 57 months (π_{1i}). The residual (*e_{it}*) represents the portion of child *i*'s VOCD score that is not predicted by his or her age. Equation (2) then calculates each child's individual growth rate (π_{1i}).¹ π_{1i} is calculated by summing the average growth rate for the sample (γ_{10}) and the extent to which each child differs from the mean growth rate (*u_{1i}*). Thus, each child has his or her own value for π_{1i} that is calculated from the average rate of change (i.e., the fixed effect γ_{10}) as well as how close or far away he or she is to that average trajectory (i.e., the random effect *u_{1i}*).

Results

The results are reported as follows: First, we report preliminary analyses and descriptive statistics of covariate predictor variables. Second, we describe relevant parameters from the random coefficient model and provide descriptive statistics for the IPSyn and VOCD growth rates. Third, we conducted a series of logistic regression models using the IPSyn and VOCD growth rates, demographic factors, and standardized assessment variables as predictor variables to determine whether a combination of these variables can predict the likelihood of stuttering recovery. Significance values were set at $\alpha = .05$ for all analyses.

Preliminary Analyses

First, we ensured that stuttering outcome was not related to the speech sample size by examining potential

¹We set the intercept—children's productive language values at study onset—to the sample average rather than letting this value randomly vary like what was done with the slope. We made this decision because preliminary analyses revealed that allowing both the intercept and slope to randomly vary was too complex to fit the data. Because our *a priori* hypothesis was that growth rates are predictive of recovery, we opted to maximize variation in this term rather than partition the variation between a random intercept and slope.

group differences in children's total utterances. It could be argued that children's language output is facilitated by improved fluency. For our sample, this does not appear to be the case. A one-way ANOVA revealed no significant differences between cohorts of children at study onset (Time 1) in number of utterances or general willingness to converse (CWS-R: $M = 298.89$, $SD = 96.29$; CWS-P: $M = 311.61$, $SD = 114.01$; CWNS: $M = 299.53$, $SD = 104.96$), $F(2,62) = 0.45$, $p = .64$. Interestingly, all children in the sample produced slightly fewer, but longer, utterances at Time 3 than at Time 1. Nevertheless, there were no statistical differences in speech sample size among groups at Time 3 (CWS-R: $M = 258.28$, $SD = 100.15$; CWS-P: $M = 224.81$, $SD = 71.63$; CWNS: $M = 238.57$, $SD = 63.49$), $F(2,64) = 0.98$, $p = .38$.

Next, we examined associations between productive language growth rates, demographic variables, and standardized assessment measures. Interestingly, IPSyn and VOCD growth rates were only marginally associated with one another ($r = .24$, $p = .08$). This suggests that they measure different linguistic constructs, as we will return to in the discussion. Maternal education was significantly and positively correlated with IPSyn growth rates ($r = .31$, $p = .03$), but not VOCD rates ($r = .15$, $p = .31$). Neither gender nor age of onset was associated with IPSyn or VOCD growth rates. Furthermore, CWS's VOCD and IPSyn growth rates were associated with many of the standardized assessment scores. IPSyn growth was positively and significantly associated with SPELT-3 ($r = .48$, $p < .001$), TACL-3 ($r = .30$, $p = .03$), and nonword repetition scores ($r = .54$, $p < .001$). VOCD growth was less strongly associated with assessment measures. VOCD growth was correlated with TACL-3 scores ($r = .27$, $p = .05$) and marginally related to SPELT-3 ($r = .24$, $p = .08$) and nonword repetition scores ($r = .25$, $p = .09$).

CWS Vary in Their Productive Language Growth Rates

Output from the random coefficient model suggested that growth in IPSyn and VOCD marginally varied between children in the sample. The PROC Mixed program in SAS, maximum likelihood estimation, was used to fit all models. As described above, the model contained a fixed effects and random effects parameter to describe the average rate of change and the extent to which the child's rate of change differed from the sample average, respectively. For the VOCD model, the fixed effect parameter indicated that the average vocabulary diversity in the sample was 81.90 VOCD units at 57 months ($SE = 1.39$), and the fixed effect linear parameter indicates that the average increase in children's vocabulary diversity was 0.23 VOCD units per month ($SE = 0.08$). For syntax, children's IPSyn scores averaged 81.69 ($SE = 0.61$) at 57 months and increased, on average, 0.22 ($SE = 0.03$) IPSyn units per month.

However, we were more interested in each child's individual growth rate, as represented by the parameters in Equation (2). Variance in the random slopes for VOCD

was 0.06 ($SE = 0.05$), which was marginally different from zero, $z = 1.35$, $p = .08$. Variance in the random slopes of IPSyn suggests that there is also marginally significant between-child variation in productive syntax (variance = 0.02, $SE = 0.02$), $z = 1.12$, $p = .10$. Thus, these marginally significant differences provide suggestive evidence that there is meaningful variation in CWS's growth rates, and the next step was to examine whether this variation would predict the likelihood of stuttering recovery.

Building a Prediction Model of Recovery Status

Logistic regression analyses were used to address the research questions by systematically exploring the predictive capacity of three groups of explanatory variables: productive language growth rates, demographic factors, and standardized assessment scores. Our model-building strategy to predict the likelihood of recovery was as follows: We first examined the effect of language growth rates on their own and then added demographic and standardized assessment factors associated with recovery status in a final prediction model. This final prediction model can reveal (a) whether language growth is a predictor over and above established correlates and (b) what the unique contribution of each variable is on the likelihood of recovery.

Is Language Growth Predictive of Stuttering Recovery?

The first two research questions were answered by examining whether children's IPSyn and VOCD growth rates predict likelihood of recovery. Of note, growth rates were rescaled from rate of change *per month* to rate of change *per year* to make each variable in the model on a more similar scale and the parameters more interpretable. With these rescaled variables, we modeled the effect of IPSyn and VOCD growth rates on recovery status in two separate logistic regression models to determine whether each linguistic domain on its own predicts eventual recovery of those with stuttering diagnosis at entry.

Table 3 indicates that IPSyn growth rates on their own were significantly and positively related to recovery status, Wald χ^2 statistic = 7.07, $p = .01$, odds ratio (OR) = 11.10. Figure 1 plots the average trajectories of CWS-P versus CWS-R, illustrating a significant difference in slopes by group. To interpret this significant finding, we calculated the predicted probability of whether or not an initially diagnosed child eventually recovered. Specifically, we use the predicted probability of recovery for a child whose IPSyn growth was at the 10th percentile of the sample and a child whose rate was at the 90th percentile for illustrative purposes. These figures indicate striking differences for children at the bottom and top deciles of the sample: For a child whose IPSyn growth rate was at the 10th percentile, the probability of recovery was 21.9%, whereas for a child at the 90th percentile, the probability of recovery was 85.3%.

We sought to determine whether growth in lexical diversity, as measured by VOCD growth rates, may also be predictive of recovery status. Table 3 reveals that the association between VOCD growth rates and recovery status

Table 3. Logistic regression results predicting recovery status from productive language growth estimates, demographic variables, and standardized assessments.

Predictors	<i>B</i>	<i>SE</i>	Wald statistic	95 CI for OR		
				Lower	OR	Upper
Syntactic complexity						
IPSyn growth	2.41	0.91	7.07**	1.88	11.10	65.39
Vocabulary diversity						
VOCD growth	0.16	0.21	0.64	0.79	1.18	1.76
Final model						
Maternal education	0.11	0.41	0.07	0.50	1.12	2.51
Nonword repetition	−0.06	0.40	0.02	0.92	0.99	1.08
BBTOP	0.11	0.04	7.29**	1.03	1.11	1.20
IPSyn growth	2.75	1.35	4.12*	1.10	15.60	221.20

Note. BBTOP = Bankson-Bernthal Test of Phonology; IPSyn = Index of Productive Syntax; VOCD = measure of vocabulary diversity (Malvern & Richards, 1997); OR = odds ratio.

* $p < .05$. ** $p < .01$.

did not reach statistical significance, Wald χ^2 statistic = 0.64, $p = .42$, OR = 1.18. Thus, growth in productive syntax but not vocabulary diversity predicts likelihood of stuttering recovery among a sample of preschool-aged children.

Is Language Growth Predictive of Recovery Over and Above Covariates?

Although IPSyn growth rates on their own predicted the likelihood of recovery, it is possible that these effects will be subsumed by established correlates of recovery status when considered together in one model. Alternatively, it is possible that IPSyn growth has unique explanatory power in predicting recovery status and will remain significant even when controlling for these factors. To test these two

Figure 1. Average Index of Productive Syntax (IPSyn) growth rate for each group of children who stutter, illustrating that, on average, children who went on to recover had steeper IPSyn growth than children whose stuttering persisted. Gray shading indicates 1 SD above and below the mean.

possibilities, we built a model with demographic, standardized assessment variables and IPSyn growth rates as predictor variables. To reduce the number of variables in this final model, we conducted two preliminary multivariate ANOVA models to determine which, if any, of these demographic and standardized assessment variables were related to recovery status. Variables that did not differ significantly as a function of recovery status were not included in the final logistic regression model.

First, group differences in three demographic variables implicated in extant literature to be related to stuttering status were examined: family socioeconomic status (as measured by maternal education), gender, and age of stuttering onset. A multivariate ANOVA suggested that these demographic variables did not significantly differ between groups (Persistent vs. Recovered), $F(3,43) = 1.34$, $p = .27$, Wilk's $\Lambda = 0.91$. Although the omnibus F statistic revealed no significant differences, maternal education was associated with IPSyn growth rates, as reported above in the Preliminary Analyses, and we thus opted to include it in our prediction model.

In the second ANOVA, we examined group differences in four verbal and nonverbal standardized assessments administered at study onset: receptive language (TACL-3), productive language (SPELT-3), phonology (BBTOP), and nonword repetition scores. There was a statistically significant difference in standardized assessment scores as a function of recovery status, $F(4,40) = 5.27$, $p = .002$, Wilk's $\Lambda = 0.66$. Post hoc comparisons using independent samples t tests revealed that CWS-R ($M = 96.68$, $SD = 12.76$) scored significantly higher on the BBTOP than CWS-P ($M = 81.80$, $SD = 12.64$), $t(46) = 3.99$, $p < .001$. Similarly, nonword repetition abilities among CWS-R ($M = 77.15$, $SD = 10.48$) were significantly higher than CWS-P ($M = 68.58$, $SD = 11.87$), $t(43) = 2.56$, $p = .01$. No significant group differences were observed in TACL-3 or SPELT-3 scores, leading us to include only BBTOP and nonword repetition scores in the final logistic regression model presented below.

A Final Prediction Model

Our third and final research question determined (a) whether IPSyn growth rates were predictive of recovery over and above covariates and (b) the predictive capacity of individual covariates when controlling for the others. This model contained the following predictor variables: maternal education, BBTOP scores, nonword repetition scores, and IPSyn growth rates. As expected, this model was considerably better at predicting recovery status compared to a model with just a constant, $\Delta\chi^2(df = 4, N = 47) = 22.07, p < .001$. Examining the individual contribution of each predictor (Table 3) reveals that IPSyn growth rates remained a unique and significant predictor of stuttering recovery even after controlling for socioeconomic status, BBTOP scores, and nonword repetition scores, Wald χ^2 statistic = 4.12, $p = .04$, OR = 15.6. Similarly, BBTOP scores remained a significant and unique predictor when holding socioeconomic status, IPSyn growth rates, and nonword repetition scores constant, Wald χ^2 statistic = 7.29, $p = .01$, OR = 1.11. The unique contribution of nonword repetition scores and maternal education² did not reach statistical significance.

For this model, we report sensitivity and specificity indices to describe the success of the model at correctly classifying children as either recovered or persisted. In this case, sensitivity refers to the ability of the model to correctly predict which children eventually recovered, whereas specificity refers to the ability of the model to correctly predict which children persisted. Overall, the model yielded relatively precise sensitivity and specificity estimates, correctly classifying 88.5% of CWS as recovered and 73.7% of children as persisting. These indices resulted in an overall classification accuracy of 82.2%. For illustrative purposes, we compared this percentage to the classification accuracy of a baseline model that assumes every participant recovers, which would have a classification accuracy of approximately 58%: $(N_{\text{recovered}}/N_{\text{participants}}) \times 100$. Thus, our iterative model-building process revealed that a combination of demographic, standardized assessment and productive language growth factors best predicted a child's likelihood of stuttering recovery during the preschool period.

Discussion

This exploration of growth patterns from spontaneous productive language samples in children with either persistent or recovered stuttering provides grist for closer examination of a number of potential indices of risk for persistent stuttering. Our study also introduces a new approach for capturing developmental change by monitoring language growth trajectories as well as original profiles in CWS to inform both theoretical and clinical approaches to stuttering in young children. Our initial examination of growth rates revealed that there is meaningful variation in

productive language growth with the larger cohort of CWS that was able to inform which children are more and less at risk to persistently stutter.

From reports of this particular cohort both in this article and elsewhere (e.g., Spencer & Weber-Fox, 2014), as well as the lively ongoing debate about relative language strength in CWS, there is abundant suggestive evidence that CWS present with varying levels of language proficiency. This should not be surprising: We know that language growth varies among the broader population of CWNS (Hoff, 2006). However, our method of using growth rates to capture maturation of expressive grammatical and lexical profiles provides more detail on the nature of this variation during the preschool period.

In particular, one major new finding is that an IPSyn in conversational speech samples from children whose stuttering remitted or persisted predicted recovery likelihood over and above demographic and standardized language assessments, each of which also yielded potential indicators of risk for stuttering persistence. In simple terms, children with the highest profiles of expressive syntax were roughly four times more likely to recover from stuttering compared to children whose samples fell at the bottom of the distribution, well beyond a standard deviation from the mean. We will leave for later in this discussion what such low scores for more complex syntax might mean for changes in intervention for the CWS showing such profiles and address first what this study's findings may tell us about the long-term concern about finding child-specific indices of the risk for stuttering persistence.

First, it appears that growth may be a more informative and sensitive measure for understanding the possible connection between language and communication disorders compared to measures taken at one point in development. This was true of our use of intake language samples for the cohort of children followed in this series of studies. It was not so much how children's expressive language samples looked when first assessed as much as how they changed over time that eventually distinguished persistent from recovered children, particularly when partnered with results of rather extensive speech-language and phonological awareness testing upon study entry. We want to emphasize that we built starting language status into our model of recovery as well, because it is well known that the language skills you demonstrate at any given point in time are necessarily bootstrapped by whatever prior skills you have already mastered. Thus, our model incorporated both static measures of skill at intake as well as growth in skills over time. We suspect that not every clinician gathers such extensive information upon referral for a stuttering problem in a preschool child; pending validation by additional research, our current studies suggest that perhaps such deeper assessment may be of both clinical as well as theoretical importance.

Next, we turn to the question of why IPSyn growth rates appear to distinguish between recovered and persistent children. We note first that more advanced profiles of expressive syntax identifying children as likely to recover

²Note that the results of the logistic regression model remain the same when including or excluding maternal education as a predictor.

were *not* the by-product of children simply saying more because they were more fluent. The groups of children we observed did not show significant differences in expressive language sample size or measures of utterance length over the time periods that we analyzed. We venture a hypothesis that will require further examination to verify. IPSyn distinguishes itself from some other measures of expressive language use, such as the Developmental Sentence Score (DSS; Lee & Canter, 1971), in that it caps the score achievable by a child after two observations of each targeted construction. Thus, a child who employs a more diverse array of constituent and phrasal forms in his or her conversational language use will continue to accumulate points, whereas a child who uses a more constrained set of what we might call “formulas” for assembling sentences will wind up showing relatively little growth, at least on this index. A recent study noted a tendency for IPSyn scores to “plateau” during the preschool years in typically developing children; they appear to do this by basically “ceiling out” at the highest values potentially achievable using IPSyn’s scoring protocol (Bernstein Ratner & MacWhinney, 2016), despite the well-attested ability of this measure to distinguish between such children and those with identified problems in language development. One possibility is that children with relatively poorer expressive language skills plateau earlier and/or at lower average levels of diversity in the types of sentence forms that they produce. This hypothesis requires additional examination by breaking down exactly how the children’s scores were achieved. Because of its extremely time-consuming nature, IPSyn reports in the published literature for large cohorts of children are typically machine-assisted, as ours was, using CLAN (MacWhinney, 2000). To answer whether our persistent children achieved lower scores over time because they repeatedly used structures that had already been awarded the maximum allowed points will require analysis of samples on an utterance-by-utterance basis. Our finding suggests that such labor might yield informative results.

Why did lexical diversity growth (as modeled by VOCD) not predict recovery status? We expected, frankly, to see some evidence that VOCD played some role in persistence and recovery, given other investigations that showed relative depression of expressive vocabulary diversity in a different cohort of CWS and fluent peers when first referred to the clinic (e.g., Silverman & Bernstein Ratner, 2002; Wagovich & Bernstein Ratner, 2007). Our follow-up analyses indicated that all CWS’s VOCD values grew relatively steeply when compared to CWNS after controlling for expressive vocabulary diversity at first visit. So, growth rates in VOCD were not able to distinguish CWS-R from CWS-P. It could easily be that another measure of lexical diversity may be better at distinguishing groups, and there is continuing dispute about what best measures to employ in analysis of samples from individuals both with and without suspected expressive language delay or disorder (e.g., Bernstein Ratner & MacWhinney, 2016; Fergadiotis et al., 2013). There was a uniform lack of correlation between VOCD values and most variables we tracked in the study, suggesting that IPSyn taps a quite separable

construct in gauging syntactic growth, unlike a measure such as DSS, which specifically “points” variations in function word use (e.g., awarding higher points to some *Wh*- words than others, different types of negative forms, modals, etc., given their developmental emergence). We could not use DSS scores appropriately in the current analyses because each child’s utterance (and there were over 50,000 of them) would have had to have been individually audited to award/withhold the mandatory “sentence point” for correctness.

This fact and our earlier observation about understanding of how CWS-P’s IPSyn scores distinguish them from their recovered peers underscore the difficult trade-off between machine-assisted analysis of large corpora obtained during research studies and the immediate understanding of *how* language sample analysis profiles emerge and change over time. What this preliminary investigation suggests is that language sample analysis profiles at onset and their growth over time may inform risk factors for recovery and persistence of early stuttering; it will require much more study to understand *why* as well as *how* these profiles inform this prognosis.

As one of the current authors has noted, intervention for childhood stuttering does not occur in a linguistic vacuum, and one cannot even begin to work on stuttering without talking, hopefully in natural connected discourse (Hall et al., 2007). Certainly, it is well attested, even in straightforwardly behavioral approaches to stuttering treatment, such as the Lidcombe Program, that syntactic complexity plays a significant role in challenging fluency in CWS (Packman et al., 2016, p. 8). Thus, no matter what approach to treating early stuttering is selected as an option, it is advisable to consider scaffolding the child’s tasks or expectations in less structured environments to gradually support the concurrent demonstration of both fluency as well as skill in producing developmentally more challenging utterances. We do note, as in some earlier work (Bernstein, 1981), that, to the extent that using developmentally easier language targets might help children achieve greater fluency, it is not a very successful strategy either long term or at all for the children in the study; by project end, children with persistent stuttering were both still stuttering and had less developmentally advanced conversational language profiles.

Our study is limited by several factors that warrant discussion. First, we did not include all variables implicated in stuttering persistence and recovery, which potentially limits how generalizable the final prediction model is for future use in identifying children who may be at risk for persisting. For instance, motor abilities have been shown to underlie stuttering onset and recovery across numerous reports (Ambrose et al., 2015; Walsh et al., 2015), though we did not include this factor in our analysis. Instead, we chose to focus on the question of how language may be associated with recovery, given the current theoretical debate in the field. It was therefore beyond the scope of the research questions to include additional factors such as motor abilities. A second limitation involves an inherent design issue related to the larger Purdue Stuttering Project.

Given that children entered the study 1–2 years after their stuttering began, it was necessary to model growth rates for when children were 57 months old, the average age at which children entered the study. Admittedly, we would have chosen to model rates closer to stuttering onset, but that was not possible given the data available for this particular sample. We should note that at age 57 months both CWS-P and CWS-R were stuttering, as reported by clinical judgment. Thus, the fact that we detected a significant difference in these growth rates that far from stuttering onset suggests that a future study that can model language growth closer to stuttering onset may have even more power to differentiate groups.

In summary, our work provides only an initial exploration of the potential benefits of tracking children's language skills in CWS over the early years of the disorder. It is possible that our current finding, that expressive grammatical maturity (as indexed via language sample analysis using IPSyn as well as standardized test measures), both at onset and over time, might provide timely prognostic information regarding the likelihood that an early stuttering problem is more or less likely to resolve itself. At the same time, should results of such testing and monitoring yield data suggestive of greater risk for persistence, we may be able to develop more nuanced therapeutic advisement to help higher-risk children to acquire greater expressive language skill as well as improved fluency.

Acknowledgments

This work was funded by NIH-NIDCD DC000559, awarded to Christine M. Weber. We are grateful to Patricia Zebrowski and her research team at the University of Iowa for obtaining some of the language samples utilized in this study. Thank you to the student research assistants of the Purdue Stuttering Project for assisting in the initial language sample transcriptions.

References

- Ambrose, N., Cox, N., & Yairi, E. (1997). The genetic basis of persistence and recovery in stuttering. *Journal of Speech, Language, and Hearing Research*, 40(3), 567–580.
- Ambrose, N. G., Yairi, E., Loucks, T. M., Seery, C. H., & Throneburg, R. (2015). Relation of motor, linguistic and temperament factors in epidemiologic subtypes of persistent and recovered stuttering: Initial findings. *Journal of Fluency Disorders*, 45, 4512–4526.
- Bankson, N. W., & Bernthal, J. E. (1999). *Bankson-Bernthal Test of Phonology*. Austin, TX: Pro-Ed.
- Bauerly, K., & Gottwald, S. (2009). The dynamic relationship of sentence complexity, childhood stuttering, and grammatical development. *Contemporary Issues in Communication Science & Disorders*, 3, 614–625.
- Bernstein, N. (1981). Are there constraints on childhood dysfluency? *Journal of Fluency Disorders*, 6, 341–350.
- Bernstein Ratner, N. (1998). Linguistic and perceptual characteristics of children at stuttering onset. In E. Charles Healey & H. F. M. Peters (Eds.) *Proceedings of the Second World Congress on Fluency Disorders* (pp. 3–6). Nijmegen, the Netherlands: Nijmegen University Press.
- Bernstein Ratner, N., & MacWhinney, B. (2016). Your laptop to the rescue: Using the CHILDES archive and CLAN utilities to improve child language sample analysis. *Seminars in Speech and Language*, 37(2), 74–84.
- Bernstein Ratner, N., & Sih, C. C. (1987). The effects of gradual increases in sentence length and complexity on children's dysfluency. *Journal of Speech and Hearing Disorders*, 52(3), 278–287.
- Bloodstein, O. (1995). *A handbook on stuttering* (5th ed.). San Diego, CA: Singular.
- Bloodstein, O. (2006). Some empirical observations about early stuttering: A possible link to language development. *Journal of Communication Disorders*, 39(3), 185–191.
- Bloodstein, O., & Bernstein Ratner, N. (2008). *A handbook on stuttering* (6th ed.). Clifton Park, NY: Thomson Delmar.
- Boyle, C., Boulet, S., Schieve, L., Cohen, R., Blumberg, S., Yeargin-Allsopp, M., . . . Kogan, M. D. (2011). Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics*, 128, 385–395.
- Bricker-Katz, G., Lincoln, M., & McCabe, P. (2009). A life-time of stuttering: How emotional reactions to stuttering impact activities and participation in older people. *Disability and Rehabilitation*, 31(21), 1742–1752.
- Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.
- Carrow-Woolfolk, E. (1999). *Test for Auditory Comprehension of Language—Third Edition*. Austin, TX: Pro-Ed.
- Chang, S., Erickson, K. I., Ambrose, N. G., Hasegawa-Johnson, M. A., & Ludlow, C. L. (2008). Brain anatomy differences in childhood stuttering. *Neuroimage*, 39(3), 1333–1344.
- Clark, C. E., Conture, E. G., Walden, T. A., & Lambert, W. E. (2015). Speech-language dissociations, distractibility, and childhood stuttering. *American Journal of Speech-Language Pathology*, 24(3), 480–503.
- Cook, S., Donlan, C., & Howell, P. (2013). Stuttering severity, psychosocial impact and lexical diversity as predictors of outcome for treatment of stuttering. *Journal of Fluency Disorders*, 38(2), 124–133.
- Dawson, J. I., Stout, C. E., & Eyer, J. A. (2003). *SPILT-3: Structured Photographic Expressive Language Test*. DeKalb, IL: Janelle.
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136–1146.
- Durán, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics*, 25(2), 220–242.
- Dworzynski, K., Remington, A., Rijdsdijk, F., Howell, P., & Plomin, R. (2007). Genetic etiology in cases of recovered and persistent stuttering in an unselected, longitudinal sample of young twins. *American Journal of Speech-Language Pathology*, 16(2), 169–178.
- Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *American Journal of Speech-Language Pathology*, 22(2), S397–S408.
- Gaines, N. D., Runyan, C. M., & Meyers, S. C. (1991). A comparison of young stutterers' fluent versus stuttered utterances on measures of length and complexity. *Journal of Speech and Hearing Research*, 34, 3437–4742.
- Hall, N., Wagovich, S., & Bernstein Ratner, N. (2007). Language considerations in childhood stuttering. In E. Conture & R. Curlee (Eds.), *Stuttering and related disorders of fluency* (3rd ed., pp. 153–167). New York, NY: Thieme.

- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Psychology*, 26, 55–88.
- Hollingshead, A. (1975/2011). Four factor index of social status. *Yale Journal of Sociology*, 8, 11–19. Retrieved from http://s3.amazonaws.com/academia.edu.documents/30754699/yjs_fall_2011.pdf
- Hollister, J., Van Horne, A. O., & Zebrowski, P. (2017). The relationship between grammatical development and disfluencies in preschool children who stutter and those who recover. *American Journal of Speech-Language Pathology*, 26(1), 44–56.
- Lai, S. A., & Schwanenflugel, P. J. (2016). Validating the use of D for measuring lexical diversity in low-income Kindergarten children. *Language, Speech, and Hearing Services in Schools*, 47(3), 225–235.
- Lee, L. L., & Canter, S. M. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders*, 36, 315–340.
- MacPherson, M., & Smith, A. (2013). Influences of sentence length and syntactic complexity on the speech motor control of children who stutter. *Journal of Speech, Language, and Hearing Research*, 56(1), 89–102.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58–71). Bristol, England: Multilingual Matters.
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. New York, NY: Palgrave Macmillan.
- McAllister, J., Collier, J., & Shepstone, L. (2013). The impact of adolescent stuttering and other speech problems on psychological well-being in adulthood: Evidence from a birth cohort study. *International Journal of Language & Communication Disorders*, 48(4), 458–468.
- Miller, J. F., & Chapman, R. S. (1984). *SALT: Systematic Analysis of Language Transcripts*. Madison, WI: University of Wisconsin–Madison.
- Mohan, R., & Weber, C. (2015). Neural systems mediating processing of sound units of language distinguish recovery versus persistence in stuttering. *Journal of Neurodevelopmental Disorders*, 7(1), 1–13.
- Nippold, M. A. (1990). Concomitant speech and language disorders in stuttering children: A critique of the literature. *Journal of Speech and Hearing Disorders*, 55(1), 51–60.
- Nippold, M. A. (2012). Stuttering and language ability in children: Questioning the connection. *American Journal of Speech-Language Pathology*, 21(3), 183–196.
- Nippold, M. A., Schwarz, I. E., & Jescheniak, J. D. (1991). Narrative ability in school-age stuttering boys: A preliminary investigation. *Journal of Fluency Disorders*, 16(5–6), 289–308.
- Ntourou, K., Conture, E. G., & Lipsey, M. W. (2011). Language abilities of children who stutter: A meta-analytical review. *American Journal of Speech-Language Pathology*, 20(3), 163–179.
- Owen, A., & Leonard, L. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: Application of D. *Journal of Speech, Language, and Hearing Research*, 45(5), 927–937.
- Packman, A., Onslow, M., Webber, M., Harrison, E., Arnott, S., Bridgman, K., . . . Lloyd, W. (2016). *The Lidcombe Program treatment guide*. Retrieved from http://sydney.edu.au/health-sciences/asrc/docs/lp_treatment_guide_2016.pdf
- Paden, E. P., & Yairi, E. (1996). Phonological characteristics of children whose stuttering persisted or recovered. *Journal of Speech and Hearing Research*, 39(5), 981–990.
- Paden, E., Yairi, E., & Ambrose, N. (1999). Early childhood stuttering II: Initial status of phonological abilities. *Journal of Speech, Language, and Hearing Research*, 42(5), 1113–1124.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: Sage.
- Rees, D. I., & Sabia, J. J. (2014). The kid's speech: The effect of stuttering on human capital acquisition. *Economics of Education Review*, 38, 76–88.
- Reilly, S., Onslow, M., Packman, A., Cini, E., Conway, L., Ukoumunne, O. C., . . . Wake, M. (2013). Natural history of stuttering to 4 years of age: A prospective community-based study. *Pediatrics*, 132(3), 460–467.
- Rescorla, L., & Turner, H. L. (2015). Morphology and syntax in late talkers at age 5. *Journal of Speech, Language, and Hearing Research*, 58(2), 434–444.
- Richels, C., Johnson, K., Walden, T., & Conture, E. (2013). Socio-economic status, parental education, vocabulary and language skills of children who stutter. *Journal of Communication Disorders*, 46(4), 361–374.
- Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, 83(2), 508–525.
- Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics*, 11(01), 1–22.
- Silverman, S., & Bernstein Ratner, N. (2002). Measuring lexical diversity in children who stutter: Application of vocd. *Journal of Fluency Disorders*, 27(4), 1–16.
- Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modelling change and event occurrence*. New York, NY: Oxford University Press.
- Spencer, C., & Weber-Fox, C. (2014). Preschool speech articulation and nonword repetition abilities may help predict eventual recovery or persistence of stuttering. *Journal of Fluency Disorders*, 41, 4132–4146.
- Usler, E., & Weber-Fox, C. (2015). Neurodevelopment for syntactic processing distinguishes childhood stuttering recovery versus persistence. *Journal of Neurodevelopmental Disorders*, 7(1), 1–22.
- Wagovich, S., & Bernstein Ratner, N. (2007). Frequency of verb use in young children who stutter. *Journal of Fluency Disorders*, 32, 79–94.
- Walsh, B., Mettel, K. M., & Smith, A. (2015). Speech motor planning and execution deficits in early childhood stuttering. *Journal of Neurodevelopmental Disorders*, 7(1), 1–12.
- Watkins, R., & Yairi, E. (1997). Language production abilities of children whose stuttering persisted or recovered. *Journal of Speech, Language, and Hearing Research*, 40(2), 385–399.
- Watkins, R., Yairi, E., & Ambrose, N. (1999). Early childhood stuttering III: Initial status of expressive language abilities. *Journal of Speech, Language, and Hearing Research*, 42(5), 1125–1135.
- Watson, J. B., Byrd, C. T., & Carlo, E. J. (2011). Effects of length, complexity, and grammatical correctness on stuttering in Spanish-speaking preschool children. *American Journal of Speech-Language Pathology*, 20(3), 209–220.
- Watts, A., Eadie, P., Block, S., Mensah, F., & Reilly, S. (2015). Language ability of children with and without a history of stuttering: A longitudinal cohort study. *International Journal of Speech-Language Pathology*, 17(1), 86–95.

-
- Watts, A., Eadie, P., Block, S., Mensah, F., & Reilly, S. (2017). Language skills of children during the first 12 months after stuttering onset. *Journal of Fluency Disorders*, 51, 39–49.
- Weber-Fox, C., Wray, A., & Arnold, H. (2013). Early childhood stuttering and electrophysiological indices of language processing. *Journal of Fluency Disorders*, 38(2), 206–221.
- Yairi, E., & Ambrose, N. (1992a). Onset of stuttering in preschool children selected factors. *Journal of Speech, Language, and Hearing Research*, 35(4), 782–788.
- Yairi, E., & Ambrose, N. (1992b). A longitudinal study of stuttering in children: A preliminary report. *Journal of Speech, Language, and Hearing Research*, 35(4), 755–760.
- Yairi, E., & Ambrose, N. (1999). Early childhood stuttering I: Persistency and recovery rates. *Journal of Speech, Language, and Hearing Research*, 42(5), 1097–1112.
- Yairi, E., & Ambrose, N. G. (2005). *Early childhood stuttering for clinicians by clinicians*. Austin, TX, Pro-Ed.
- Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, 38(2), 66–87.
- Yairi, E., Ambrose, N. G., Paden, E. P., & Throneburg, R. N. (1996). Predictive factors of persistence and recovery: Pathways of childhood stuttering. *Journal of Communication Disorders*, 29(1), 51–77.