

ABSTRACT

Title of Thesis: A LARGE-SCALE ANALYSIS OF LEXICAL DIVERSITY IN CHILDREN WHO STUTTER

Courtney Luckman, Master of Arts, 2017

Thesis directed by: Dr. Nan Bernstein Ratner
Department of Hearing and Speech Sciences

This study compared lexical abilities in 99 pairs of children who stutter (CWS), ages 25 – 100 months and age-, gender-, and SES-matched children who do not stutter (CWNS) in spontaneous conversation and on standardized vocabulary tests. Correlations among lexical diversity measures and dissociations between receptive and expressive standard scores were also calculated. CWS demonstrated similar lexical diversity compared to CWNS in measures computed for spontaneous speech, but a highly significant difference was found between CWS and CWNS on expressive and receptive standardized vocabulary scores. Despite prior reports, CWS were no more likely to exhibit dissociations on expressive and receptive vocabulary than CWNS. There were significant correlations among three measures of lexical diversity: number of different words (NDW), vocabulary diversity (VocD), and moving-average type-token ratio (MATTR). The effect that sample size and algorithms have on validity of measuring lexical diversity is discussed. Future directions in stuttering research are suggested.

A LARGE-SCALE ANALYSIS OF LEXICAL DIVERSITY
IN CHILDREN WHO STUTTER

by

Courtney Renee Luckman

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Arts
2017

Advisory Committee:

Professor Nan Bernstein Ratner, Chair
Professor Jan Edwards
Professor Rochelle Newman

Acknowledgements

I would like to thank my committee chair and advisor, Dr. Nan Bernstein Ratner, for her time, effort, and wisdom that she devoted to this project. I would also like to thank my committee members, Dr. Rochelle Newman and Dr. Jan Edwards, for serving on my thesis committee and providing me with excellent feedback.

Thank you to my collaborators: Soo-Eun Chang, Christine Weber-Fox, Stacy Wagovich, Nancy Hall, and Haya Berman Hakim for sharing their data. I would also like to thank all of the parents and children that participated in their studies.

I would like to thank my parents for always encouraging me to follow my passion and curiosity. And lastly, to all the people who stutter, who inspired this research.

Table of Contents	
Acknowledgements	ii
Table of contents	iii
List of tables	iv
List of figures	v
1. Introduction	1
1.1 Language abilities in children who stutter	1
1.2 Lexical skills of CWS on standardized vocabulary measures	7
1.3 Measuring lexical diversity in spontaneous speech	10
1.4 Conversational lexical skills in children who stutter	16
1.5 Summary	16
1.6 Research questions and hypothesis	17
2. Methods	18
2.1 Participants	18
2.2 Language samples	21
2.3 Language testing	21
2.4 Analysis	21
2.4.a Lexical diversity in spontaneous speech	21
2.4.b Standardized tests	23
2.4.c Relationships among lexical diversity values in spontaneous speech and standardized tests scores	24
3. Results	24
3.1 Lexical diversity in spontaneous speech of CWS and CWNS	24
3.2 Standardized tests	26
3.3 Comparing lexical diversity in spontaneous speech vs. standardized tests	29
4. Conclusions	35
4.1 Lexical diversity in spontaneous speech	35
4.2 Lexical diversity on standardized tests	36
4.3 Comparing lexical diversity in spontaneous speech vs. standardized tests	39
4.4 Limitations	42
4.5 Future research	43
4.6 Summary	44
References	46

List of Tables

Table 1: Participant data	20
Table 2: Mean lexical diversity for CWS and CWNS across measures	25
Table 3: Dissociations between receptive and expressive vocabulary	28
Table 4: Correlations among lexical diversity measures	30
Table 5: Correlations between lexical diversity measures and expressive vocabulary	34

List of Figures

Figure 1: Dissociations between expressive and receptive vocabulary	4
Figure 2: Standardized test performance	28
Figure 3: Correlation between NDW and VocD	31
Figure 4: Correlation between NDW and MATTR	32
Figure 5: Correlation between VocD and MATTR	33
Figure 6: Correlation between expressive vocabulary standardized test score and TTR50	34

1. Introduction

1.1 Language abilities in children who stutter

Children who stutter may exhibit subtle weaknesses in language. Although language weakness is an unlikely cause of stuttering, various language-related factors have been proposed to influence stuttering profiles. Onset of stuttering typically begins between ages 2-5 years, during a period of rapid language growth (Hall, Wagovich, & Bernstein Ratner, 2007). Stuttering typically appears after the child has demonstrated normal speech fluency, and is often thought to coincide with advances in language development (Watkins & Yairi, 1997). The direction and magnitude of the relationship between children's fluency and language skills in children who stutter (CWS) remains unclear. Some studies report depressed language abilities (Williams, Melrose, & Woods, 1969; Westby, 1979; Ryan, 1992; Anderson & Conture, 2000; Silverman & Bernstein Ratner, 2002; Anderson, Pellowski, & Conture, 2005; Ntourou, Conture, & Lipsey, 2011; Choo, Burnham, Hicks, & Chang, 2016) while others report no differences in language abilities (Nippold, 1990; Watkins & Yairi, 1997; Nippold, 2012; Watts, Eadie, Block, Mensah, Reilly; 2017) and some report advanced language abilities (Watkins, Yairi, & Ambrose, 1999; Bonelli, Dixon, & Bernstein Ratner, 2000; Watts, Eadie, Block, Mensah, & Reilly, 2015). Although there appear to be subtle differences in language abilities between children who stutter and fluent peers, most CWS in published studies still fall within normal limits and cannot be classified as being clinically delayed or as having a language disorder (Ntourou et al., 2011).

To reconcile these findings, several researchers (Anderson et al., 2005; Coulter, Anderson, & Conture, 2009; Choo et al., 2016) have examined dissociations among sets

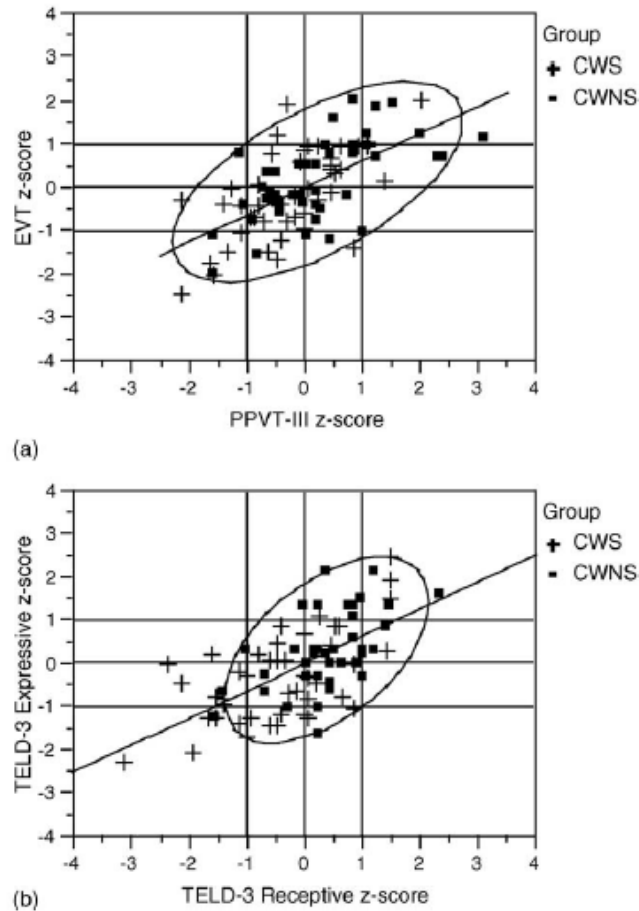
of language skills that may help account for discrepancies in language abilities of children who stutter. Instead of examining if children who stutter perform above or below fluent peers on standardized language tests, it might be more informative to look at the presence and degree of correspondence among an individual's component speech and language skills (Anderson et al., 2005). In other words, do children who stutter show mismatches between areas of strength and weakness in subsystems that support language use and comprehension? Potentially, dissociations between linguistic abilities of children who stutter may precipitate breakdowns in speech fluency suggesting trade-offs in speech and language production, or a linguistic version of "demands and capacities" stress (Watkins & Yairi, 1997; Hall, 2004; Anderson et al., 2005). In an attempt to restore these language encoding imbalances, there may be a reduction in resources that are available for fluent speech (Anderson et al., 2005; Coulter et al., 2009). If one language skill is weaker than another, different components may arrive at a central language integrator at different times and thus result in a mistiming of end-stage processes necessary for speech-motor production (Perkins, Kent, & Curlee, 1991). It is additionally possible, given this framework, that children who stutter may exhibit subtle weaknesses across various domains that may not be the same across all children who stutter. There may not be a consistently weak linguistic domain in all children who stutter.

Some research has supported the theory of dissociations among linguistic domains in CWS. In a preliminary study, Anderson and Conture (2000) examined language abilities of 20 CWS and 20 CWNS through standardized tests of receptive and expressive language: *Peabody Picture Vocabulary Test-III (PPVT-III)* (Dunn & Dunn, 1997) and *Test of Early Language Development-2 (TELD-2)* (Hresko, Reid & Hamill, 1991).

Children who stutter exhibited a greater difference between overall measures of language abilities (*TELD-2*) and receptive vocabulary (*PPVT-III*) than their fluent peers. The authors suggested that in children who stutter, semantic development (and thus, potentially, lexical retrieval) might trail syntactic development, creating an imbalance that disrupts ongoing speech-language production.

Subsequently, Anderson et al. (2005) examined dissociations in speech and language skills in 45 CWS and 45 CWNS matched on age, gender, race, and socioeconomic status using the same standardized test measures as Anderson and Conture (2000), with an added test of expressive vocabulary, the *Expressive Vocabulary Test (EVT)* (Williams, 1997). The researchers calculated correlations between z-scores on standardized tests to measure dissociations, or asynchronies between various speech and language domains. Dissociations were then identified using density ellipses that displayed the degree of correlation and identified outliers. They found that there were more outliers who were CWS, rather than CWNS, signifying more dissociations between language measures in CWS than CWNS. These profiles are illustrated in Figure 1:

Figure 1: Dissociations between expressive and receptive vocabulary (Anderson et al., 2005)



The researchers took these results to suggest that:

“Perhaps it is the child’s attempt to reconcile or manage these dissociations in speech and language that contributes to disruptions in their speech-language production, which in combination with a genetic predisposition towards stuttering, or perhaps a temperamental disposition that is relatively intolerant of any such disruptions, results in the emergence of persistent stuttering.” (2005, page 242)

Coulter (2009) found results similar to Anderson and Conture (2000) and Anderson et al. (2005) using a larger sample of matched participants (85 CWS and 85 CWNS), overlapping with the 45 participants studied by Anderson et al. (2005). Using the same standardized test battery as Anderson et al. (2005), the researchers found that children who stutter were three times more likely to exhibit dissociations across speech-language domains, a similar ratio to the team's earlier studies, but not unexpected, since more than half the children were used in both prior studies. Using the same method to calculate dissociations, the researchers found that the greatest dissociations were between the domains of receptive and expressive language abilities (tested receptive language ability was significantly below expressive language). The researchers concluded that this dissociation could result in children who stutter attempting to produce language beyond their general language capacities, creating fluency-hindering demands on the speech-language processing system if the child attempted to talk within modeled speech rates. This would be somewhat analogous to a second language learner attempting to engage in conversation with more proficient speakers using the native conversational tempo.

Beyond the linguistic domain, there is also research to support dissociations among cognitive (Choo, 2016) and motor (Hollister, 2012; Smith, Goffman, Sasisekaran, & Weber-Fox, 2012; MacPherson & Smith, 2013) domains in children who stutter. CWS showed higher rates of dissociations between language and IQ, as measured by standardized tests of language (*PPVT*, *EVT*, *TOLD*, *TACL*) and cognition (*Weschler Preschool and Primary Scale of Intelligence (WPPSI-III)* and *Weschler Abbreviated Scale of Intelligence (WASI)*). In motor performance measures, research suggests

dissociations in CWS between diadochokinesis (DDK) and speech rate, as well as between expressive language and DDK.

Some skills may be developed earlier, while others appear delayed (Choo et al., 2016). In fact, this imbalance among domains of language development is thought to characterize other speech and language disorders, in which the child reaches a peak in language level where any extra demands cause a breakdown in the system (Crystal, 1987; the “bucket theory” of child language processing).

One popular model describing the development of stuttering, the Demands and Capacities Model, or DCM (Adams, 1990; Starkweather, 1987; Starkweather & Gottwald, 1990), proposes that fluency breakdowns are related to the child’s inability to successfully manage the linguistic components of producing an utterance. Incongruity between the environmental demands and individual capacities across a language domain may impact speech motor production. Starkweather (1987) introduced the idea of demands and capacities by theorizing that a child who stutters chronically lacks the capacity to meet demands for fluency. One of the stressors on fluency comes from the child’s language development. If language skills are weakened or imbalanced in children who stutter, this could produce a greater likelihood that environmental communicative demands exceed the child’s ability to produce fluent speech.

Based on where and when children tend to stutter, we can theorize that children who stutter may experience subtle weaknesses in efficient encoding and retrieval of lexical and grammatical targets (Ntourou et al., 2011). Some research demonstrates that disfluencies increase as utterance length and complexity increases (Logan & Conture, 1995, 1997; Zackheim & Conture, 2003), while others find that developmental measures

of utterance complexity, rather than length, better accounts for fluency breakdown (Bernstein Ratner & Costa Sih, 1987; Gaines, Runyan, & Meyers, 1991). Additionally, children tend to stutter on function words, short words, and words that initiate syntactic units (Bernstein, 1981; Bloodstein & Grossman, 1981), in contrast to an inverse pattern seen in most adults, where stutters tend to migrate toward longer and comparatively rare open class lexical items. It is unclear whether syntactic formulation or lexical retrieval best accounts for this profile. The subtle weaknesses in language that CWS appear to display may be due to underdeveloped lexical abilities that make it difficult to quickly and correctly place items in a syntactic frame (Anderson & Conture, 2000). CWS may attempt utterances of varying length and complexity but due to their underdeveloped lexical abilities, this may exceed their ability to produce fluent utterances (Gaines et al., 1991).

1.2 Lexical skills of CWS on standardized vocabulary measures

While some research points to depressed vocabulary in children who stutter, some suggest advanced vocabulary, while others report no difference. This could be due to the wide range of vocabulary diversity measures used. Some studies used standardized measures such as the *PPVT*, *EVT*, and *CELF* (e.g., Coulter et al., 2009), while others used conversational analysis to calculate measures such as NDW, TTR, lexical rarity, and VocD (e.g., Watkins & Yairi, 1997). Additionally, some studies report having a control group (e.g., Anderson et al., 2005) while others do not (e.g., Bonelli et al., 2000). Some studies had inclusion requirements that participants score within normal limits on language tests (e.g., Coulter et al., 2009), while others had no such criteria. Some studies' participants were matched for a variety of factors, while others failed to match on

potentially important variables such as age, gender, and socioeconomic status (e.g., Anderson et al., 2000).

The majority of studies measuring vocabulary diversity through standardized assessments used the *Peabody Picture Vocabulary Test (PPVT)* (Dunn & Dunn, 1997) and the *Expressive Vocabulary Test (EVT)* (Williams, 1997). The *PPVT* is a measure of single word receptive vocabulary and is normed for ages 2;6 – 90+years. The examiner presents a series of pictures to the child and asks the child to point to the corresponding picture that best describes the word's meaning. The *EVT* is a measure of single word expressive vocabulary and is also normed on ages 2;6 – 90+years. The examiner presents a picture and reads a stimulus question from the record form. The child must respond with one word that provides a label for the picture.

Several researchers have found differences on these tests for children who stutter versus their age-matched peers. For receptive vocabulary, Westby (1979) found that CWS in kindergarten and first grade performed significantly more poorly than their peers on the *PPVT*. Anderson and Conture (2000), testing 20 pairs of CWS/CWNS between the ages of 3 and 5, found that CWS scored significantly lower than controls on the *PPVT*.

For expressive vocabulary, Silverman and Bernstein Ratner (2002) examined 15 children who stutter, within 3 months of stuttering onset, and 15 normally fluent peers matched on age, gender, and maternal level of education. They found that the CWS scored significantly more poorly on the *EVT* than their fluent peers.

Some studies have found differences in both expressive and receptive vocabulary among their sample. Using a different measure of receptive and expressive vocabulary and a slightly older population (children in grade six), Williams et al. (1969) found a

significant difference on the vocabulary section of the *Iowa Tests of Basic Skills* between 100 children who stuttered and 300 children who did not stutter. This subtest examines general language ability. In the first part, the child hears a word and chooses the corresponding picture (out of 3), similar to the *PPVT*, but with fewer foils. In the next part, the child must define a word in the context of a sentence. In the last part, the word is presented in a short phrase or sentence and children have to select a synonym that has the same meaning as the target word. Even though differences were apparent between the two groups on this test, the range of scores of CWS were quite varied, indicating that lexical differences may not be present in all children who stutter. Coulter et al. (2009) used a large sample of 85 CWS and 85 CWNS between the ages of 3 and 6, matched on age, gender, and race. The CWNS were required to receive a standard score of 85 or higher on each of the five standardized speech-language subtests, whereas the CWS were allowed to freely vary in their scores. CWS scored significantly lower on the *PPVT* and *EVT* than did CWNS. Choo et al. (2016) examined 66 children who stutter and 53 fluent peers ages 3 to 10 matched by group average age and group average socioeconomic status. All children underwent screening to ensure normal speech and language (aside from stuttering). The researchers found that CWS scored significantly lower on the *PPVT* and *EVT* than did CWNS.

Although across these studies CWS performed worse on vocabulary measures than did CWNS, they still scored within the normal range on all measures. This suggests that, as a group, CWS are not language-impaired, but rather may have subtle weaknesses in lexical knowledge or retrieval.

Other researchers have found no differences on these same tests for children who stutter versus their age-matched peers. Using pairs matched on age, gender, and SES, Ryan (1992), Silverman & Ratner (2002), and Anderson et al. (2005) found that CWS performed similarly to CWNS on the *PPVT*. Millager (2014) examined 40 preschool age CWS and 46 CWNS and found no significant differences between groups on the *PPVT* or the *EVT*. Participants in this study had to score above the 16th percentile on these tests to be included in this study.

In an attempt to remediate conflicting findings, Nippold (1990) constructed a critical review of studies using both standardized language tests and spontaneous language sample measures and found little support for any substantive differences in vocabulary between children who stutter and children who do not stutter. In contrast, Ntourou et al. (2011) performed a meta-analysis of 22 studies conducted after 1990 with participants between 2 and 8 years of age who had all been assessed using norm-referenced language tests or language sampling tasks. They found that CWS scored significantly more poorly on norm-referenced tests of both receptive and expressive vocabulary. The authors argued for consistent differences in language abilities, rather than dissociations between linguistic skill domains, among children who stutter. These differences may not be evident if researchers do not use a control group, but rather compare scores to pre-established norms (e.g., comparing a group's TTR to the published norms) (Watkins et al., 1999; Bonelli et al., 2000), which may not be appropriate to their specific sample.

1.3 Measuring lexical diversity in spontaneous speech

In addition to standardized measures, several researchers have used spontaneous language analysis to examine lexical diversity in children who stutter. A number of different ways to calculate lexical diversity have been proposed. The oldest measure to be used in assessing children's language development is type-token ratio (Templin, 1957). TTR was defined by Johnson (1944) as the ratio of different words (types) to total words (tokens). TTR ranges from 0 – 1, with higher values representing greater lexical diversity. However, due to the influence of closed class morphology in creating grammatically correct utterances in any language, typical TTR values rarely exceed .5 in typical conversation, as noted by Templin (1957).

Templin (1957) calculated TTR on 50-utterance samples from 480 children ages 3 – 8 years. Half of the sample was male and half was female. There were 60 children each for ages 3;5, 4;0, 4;5, 5;0, 6;0, 7;0, and 8;0 (years; months), 144 children came from an upper SES community and 336 came from a lower SES sample. Using 50-utterance samples, Templin found a TTR of approximately .50 that did not vary across age, gender, or socioeconomic status. Although Templin never used the term TTR, she described this ratio as “approximately one different word for slightly over every two words uttered” during spoken interactions.

Using Templin's data, Miller (1981) pooled numbers to derive a mean for each of the eight age groups of children. He concluded that the consistency of this measure across age groups makes it valuable as a clinical tool. He proposed that if the child's TTR is below 0.5, it may indicate that the child is language-impaired.

Most researchers do not find TTR to be a reliable or valid measure of lexical skills in children (Hess, Ritchie, & Landry, 1984; Hess, Haug, & Landry, 1989; Richards,

1987; Hess et al., 1989; Malvern, Richards, Chipere, & Duran, 2004; Koizumi & In'nami, 2012; Bernstein Ratner & MacWhinney, 2016). First, TTR in children's language samples do not tend to align with the same children's performance on standardized vocabulary comprehension or production measures. In their sample of children, Hess et al. (1984) found that TTR was not correlated with scores on the *PPVT*, the oral and picture subtests of the *Test of Language Development (TOLD)*, or the *Test of Written Language*. Second, TTR does not appear to reflect children's vocabulary growth as they age. The most recent example of this finding is for over 600 children in the CHILDES Archive followed by Bernstein Ratner & MacWhinney (2016). TTR should be higher for an 8-year old than a 3 year old, but Templin's data do not reflect this, nor do the CHILDES data, which vary rather wildly across the ages 2-6 years. Instead, as a child ages, he/she produces longer utterances containing more replicated grammatical free morphemes, which is negatively associated with the number of types. Additionally, as children's conversations become longer and more detailed, it becomes more difficult to avoid repetition. A chance of a new "type" being used decreases.

TTR also shows methodological weakness. Despite the use of a proportion to compute TTR, there is still a significant confound of length. Larger samples tend to yield lower TTR values. Therefore, TTR samples need to be standardized, or compared to samples of the same length. Using language samples of varying sizes from children 9 through 12 years of age (a slightly higher age range than Templin), Hess et al. (1989) found significant differences in the mean TTR on samples of different lengths. More recent research also suggests that length influences TTR measurements (Tommerdahl & Kilpatrick, 2013; Tommerdahl & Kilpatrick, 2014).

Standardizing language samples for use of TTR becomes problematic even when comparing samples of the same length. Watkins, Kelly, Harbers, & Hollins (1995) found similar TTRs between two children who both produced the same number of utterances, even though one produced significantly more total words. Using samples of the same length (100 words), TTR was found not to distinguish specifically language-impaired (SLI) children from typically- developing children, even though standardized tests did (Watkins et al., 1995). Finally, using subsets of utterances may not adequately represent a child's complete language ability and may unnecessarily waste data (Hess et al., 1989; McKee, Malvern, & Richards, 2000).

To avoid length or sample effects, number of different words in a 100-word sample (abbreviated NDW) (Miller, 1991; Klee, 1992) is frequently used due to its simplicity and ability to address the range of a child's vocabulary in a standard sized sample (Malvern et al., 2004). Klee (1992) examined children between 24 and 50 months and found that NDW was able to discriminate between typically-developing children and those with specific language impairment. He also found that NDW was strongly correlated with age.

However, Malvern et al. (2004) notes several additional problems with NDW. As with TTR, the number of different words in a certain sample depends on sample size. Therefore, it's important to use standardization. However, it is difficult to ascertain when exactly to cut the sample to maximize the informativeness of the measure. They found NDW to be significantly different using the whole sample, the first 100 words, the first 50 words, a random selection of 25 words, and a sequence of 25 contiguous words.

Additionally, Miller (1991), examining 192 typically-developing children, found that NDW correlated with MLU, which might confound utterance standardization.

In order to overcome these problems with TTR and NDW, Gerald McKee (2000) developed VocD. VocD, now a part of the CLAN software utilities (MacWhinney, 2000) uses a mathematical algorithm to calculate vocabulary diversity. It first samples random groups of words from a transcript 100 times to produce a curve of the TTR against tokens. Then, it finds a best fit between this curve and theoretical curves calculated by the model by computing the probability of new vocabulary being introduced into progressively longer samples. Final values range from 10 to 100, with higher values indicating greater diversity.

VocD is significantly less impacted by sample length than TTR (Malvern et al., 2004; McCarthy & Jarvis 2007, 2010; Koizumi & In'nami, 2012). McCarthy (2007) tested 14 lexical diversity measures on written texts and spoken samples. The measures included those frequently used in previous studies, as well as several measures that correct for sample size and frequency of types such as root TTR (RTTR) (Guirand, 1960), corrected TTR (CTTR) (Carroll, 1964), Uber (U) (Dugast, 1978), Somer's S (SS) (Somers, 1966), and Rubet's K (RK) (Dugast, 1979). All correlated with text length, but VocD had one of the lowest correlations, 0.22, suggesting less influence of sample size on the estimate of diversity. He also found texts between 100 and 400 tokens to be comparable in VocD profiles. Similarly, Malvern et al. (2004) examined the files of the New England children in the CHILDES database (Dale, Bates, Reznick, & Morriset, 1989; Snow, 1989), which consists of 38 children ranging in age from 27 to 33 months

from 17 low SES and 21 high SES families. They found that VocD scores did not change based on the number of tokens in the children's samples.

McKee notes that VocD has three advantages: it is not a function of the number of words in the sample, it uses all of the data available, and it takes into account the variability of tokens in a sample. It has been found to correlate with standardized measures of expressive vocabulary in children, while TTR values do not (Silverman & Bernstein Ratner, 2002).

Some have found, however, that VocD is not reliable on very small (Owen & Leonard, 2002; Koizumi & In'nami, 2012) or very large (McCarthy & Jarvis, 2010; Fergadiotis, Wright, & Green, 2015) samples. Most lexical diversity measures caution against use on samples under 200 tokens. On samples of greater length, VocD can be a better alternative. Very long samples may not yield a consistent VocD due to the introduction of new themes, which may cause VocD to be less reliable across repeated trials. These problems may be remediated, however, with a standardized set of toys or activities that are used by participants.

An alternative to VocD, MATTR (Moving-Average Type-Token Ratio), has been shown to be valid and reliable, even on large texts (Covington & McFall, 2010; Fergadiotis, Wright, & West, 2013; Fergadiotis et al., 2015). MATTR calculates the lexical diversity of a sample using a moving window that estimates TTR's for each successive window of fixed length. MATTR has been demonstrated to be even less influenced by sample size than VocD and appears to be more reliable than most other measures (Fergadiotis et al., 2013, 2015). However, this measure has been mainly used on people with aphasia, who may produce a slightly different lexical pattern.

1.4 Conversational lexical skills in children who stutter

Watkins and Yairi (1997) examined spontaneous language of 32 children who stutter. CWS were found to use an average or slightly higher number of different words (NDW) than the norms (Leadholm & Miller, 1992), with more lexical variability among the children who persisted in their stuttering, rather than recovered. However, adding 84 more children who stuttered to later analyses, Watkins et al. (1999) found no significant difference in NDW between the CWS and Leadholm and Miller (1992) norms. Also using NDW, Bonelli, Dixon, Onslow and Bernstein Ratner (2000) examined 9 children who stutter and found that all NDW values in spoken language were within expected range according to Templin's (1957) norms. In contrast, Silverman and Bernstein Ratner (2002) found that CWS did not differ on TTR, but did differ on VocD computations derived from their spontaneous language samples.

1.5 Summary

The review of the literature suggests that CWS may have subtle weaknesses in understanding and using vocabulary, albeit in the absence of a major deficit and with potentially greater amount of variability seen among children and samples, when compared to typically fluent peers. Disagreement among studies could reflect unique features of samples used in analysis. Several studies (Anderson & Conture, 2000; Anderson et al., 2005; Coulter et al., 2009) used overlapping subjects in their analysis. Additionally, the validity of some measures used to assess these differences in spontaneous language samples is still in question. Several studies used standardized measures, while others used measures on spontaneous language, but few have explored

the relationship between lexical diversity in spontaneous speech and standardized test scores (such as *PPVT* and *EVT*).

1.6 Research questions and hypothesis

The primary focus of this study is to examine lexical abilities in children who stutter using a large cohort of children who stutter close to onset and typically-fluent peers matched on age, gender, and socioeconomic status. We examined spontaneous conversational speech through language samples as well as standardized language testing, collapsing six longitudinal studies of children who stutter. The secondary focus of the study was to assess the validity of lexical diversity measures by comparing across them, and by evaluating how well they correlate with other measures of lexical skill, such as standardized vocabulary tests.

1. If children who stutter demonstrate subtle language weaknesses, it is hypothesized that they will score significantly lower than their peers on both standardized and spontaneous language measures of vocabulary diversity.
2. If children who stutter have an imbalance among speech-language domains, it is hypothesized that they will exhibit dissociations between receptive vocabulary and expressive vocabulary.
3. If NDW, VocD or MATTR is a more valid measure than TTR, it is hypothesized that they will be correlated more strongly with standardized measures.

If children who stutter demonstrate depressed lexical abilities on both standardized measures and spontaneous language measures through this tightly controlled, large-scale analysis, this could support the link between stuttering and

language and could further our understanding of models of stuttering. This also might help us redefine the nature of stuttering as not purely a speech motor disturbance.

Additionally, if spontaneous language measures such as VocD are comparable to standardized measures, language sample analysis assisted by computerized utilities can become a valuable tool in clinical evaluation and research studies.

2. Methods

2.1 Participants

The total number of participants was 198 (99 children who stutter and 99 children who do not stutter) derived by pooling data from several previous investigations (Silverman & Bernstein Ratner, 2002; Leech et al., in press; Choo et al., 2016; Wagovich & Hall, 2007; Hall et al., 2007; Johnson, Karrass, Conture, & Walden, 2009; Hakim & Bernstein Ratner, 2004). Participants from Silverman and Bernstein Ratner (2002) were recruited by flyers in pediatricians' offices in the greater Washington D.C. area. Participants from Leech et al. (in press) were recruited from Purdue University and University of Iowa as part of the Purdue Stuttering Project. Participants from Choo et al. (2016) were recruited through the Speech Neurophysiology Lab at Michigan State University as part of a longitudinal study examining neurodevelopmental correlates of stuttering. Participants from Wagovich and Hall (2007) were recruited through a University of Missouri community e-mail bulletin and from community daycares. Participants from Johnson et al. (2009) were recruited through the Vanderbilt University Developmental Stuttering Research Project. Participants from Hakim and Bernstein Ratner (2004) were recruited via recruitment of CWS and classmates from the greater Washington, D.C. area. Twenty CWNS were identified from the CHILDES archive (Ellis

Weismer, et al., 2013; Warren-Leubecker & Bohannon, 1984; Dickinson & Tabors, 2001) in order to pair-wise match a small proportion of CWS to fluent peers similar in gender, age, and SES. These CHILDES corpora were chosen because they contained typically-developing children engaged in toy play, similar to elicitation procedures for the CWS. All data were de-identified and were anonymous to the researcher.

All pairs were matched by age (within 3 months), gender, and socioeconomic status (mother's level of education). CWS had an average age of 51 months, with a range of 25 months to 100 months. CWNS had an average age of 50 months, with a range of 28 to 100 months. There were 71 male pairs and 28 female pairs. All pairs were matched on maternal education such that there was no significant difference between groups in years of maternal education. For all children examined in this study, there was no history of speech or language disorders other than stuttering. All CWS across studies met a criterion of at least 3% stutter-like disfluencies in their spontaneous language samples and ranged from very mild to very severe, as calculated by percent stuttered syllables (Silverman & Bernstein Ratner, 2002; Choo et al., 2016), parent and speech-language pathologist report (Leech et al, in press), or SSI-3 (Wagovich & Hall, 2007; Johnson et al., 2009; Hakim & Bernstein Ratner, 2004).

Some studies originally only examined lexical diversity in spontaneous language, others only examined lexical diversity on standardized vocabulary tests, and others examined both. One hundred and fifty-two participants (76 pairs of CWS and CWNS) were included in the spontaneous language sample analysis (LSA). This sample contained 22 pairs of females and 54 pairs of males with an average age of 53 months (range of 25 to 79 months). One hundred participants (50 pairs of CWS and CWNS) were

included in standardized test score analysis. This sample contained 14 pairs of females and 36 pairs of males, with an average age of 48 months (range of 29 to 100 months). Fifty-four participants (27 pairs of CWS and CWNS) were used in correlational analysis of spontaneous vocabulary measures and standardized vocabulary tests. This sample contained 7 pairs of females and 20 pairs of males with an average age of 48 months (range of 29 to 68 months). Full participant characteristics are summarized below.

Table 1: Participant Data

Corpus	Total # of pairs	# of female pairs	# of male pairs	# of pairs with CWNS from CHILDES archive	Age range (months)	# of pairs used in spont. lang analysis	# of pairs used in stand. test analysis	#of pairs used in correl. analysis
Bernstein Ratner (Silverman & Ratner, 2002)	20	2	18	8	28 – 46	20	7	7
Purdue (Leech et al., in press)	27	8	19	8	44 – 79	27	0	0
Michigan (Choo et al., 2016)	24	9	15	4	36 – 68	24	20	20
Wagovich (Wagovich & Hall, 2007; Hall et al., 2007)	5	3	2	0	25 – 44	5	0	0
Vanderbilt (Johnson et al., 2009)	15	6	9	0	39 – 65	0	15	0
Hakim (Hakim & Bernstein Ratner, 2004)	8	1	7	0	49 – 100	0	8	0
Total	99	29	70	20	25 – 100	76	50	27

2.2 Language samples

Language samples of clinician-child interactions (Choo et al., 2016; Wagovich & Hall, 2007; Hall et al., 2007), parent-child interactions (Warren-Leubecker & Bohannon, 1984; Ellis Weismer, Venker, Evans, & Moyle, 2013; Dickinson & Tabors, 2001) or clinician-child and parent-child interactions (Silverman & Bernstein Ratner, 2002; Leech et al., in press) engaged in toy play were obtained. Pairs were matched as much as possible to match for parent- or clinician-led language sample context. The conversations were recorded on videotape and audiotape. The language samples were then transcribed using CHAT (MacWhinney, 2000).

2.3 Language testing

Participants were given standardized vocabulary tests. Participants from Silverman and Bernstein Ratner (2002) completed the *Peabody Picture Vocabulary Test – Revised (PPVT-R)* (Dunn & Dunn, 1997) and *Expressive One-Word Picture Vocabulary Test – Revised (EOWPVT-R)* (Gardner, 1990). Participants from Choo et al. (2016) completed the *Peabody Picture Vocabulary Test (PPVT-4)* (Dunn & Dunn, 1997) and *Expressive Vocabulary Test (EVT-2)* (Williams, 2007). Participants from Johnson et al. (2009) completed the *Expressive Vocabulary Test (EVT)* (Williams, 1997) and *Peabody Picture Vocabulary Test (PPVT-3)* (Dunn & Dunn, 1997). Participants from Hakim and Bernstein Ratner (2004) completed the picture vocabulary subtest and oral vocabulary subtest of the *Test of Language Development-Primary (TOLD P-3)* (Hammill & Newcomer, 1997).

2.4 Analysis

2.4.a Lexical diversity in spontaneous speech

In the CLAN utilities, lexical diversity is computed on the morphological root (e.g., “cats” and “cat” are coded as the same word). NDW, TTR, VocD, and MATTR of the 76 pairs were analyzed using CLAN utilities (MacWhinney, 2000). NDW was calculated from the first 100 words in each child’s sample. TTR was calculated on roughly the middle 50 utterances of each sample, since this is found to be the most reliable way to sample (Miller, 1981; Retherford, 2000; Silverman & Bernstein Ratner, 2002). Analysis of TTR was started at the 50th utterance for all transcripts, with the exception of 3 transcripts that were started at the 5th utterance (because they had only 58 utterances), 4 transcripts that were started at the 10th utterance (because they were between 62 and 68 utterances long), 4 transcripts that were started at the 15th utterance (because they were between 71 and 78 utterances long), 11 transcripts that were started at the 25th utterance (because they were between 80 and 99 utterances long), and 4 transcripts that included the entire sample in TTR analysis (because they were less than 50 utterances in length). Thus, a total of 26 samples were analyzed starting at points other than the 50th utterance.

TTR of the entire sample for all transcripts was also calculated as a comparison to TTR50. VocD was calculated on the entire sample, since it has been found that there is no difference in VocD whether it is computed using a truncated sample of utterances or using the entire sample (McKee, 2000; Silverman & Bernstein Ratner, 2002). MATTR was calculated using a successive window of 50 words, because using this moving window has been demonstrated to be the most reliable and valid indicator of lexical diversity (Fergadiotis, 2011).

There were 2 CWNS who were not included in the NDW analysis due to samples containing fewer than 100 words.

Analysis:

For NDW, an independent samples *t* test was applied with group membership as the independent variable and NDW as the dependent variable. For VocD, an independent samples *t* test was applied with group membership as the independent variable and VocD as the dependent variable. For TTR50, a Mann-Whitney U test was applied with group membership as the independent variable and TTR50 as the dependent variable. For TTR, a Mann-Whitney U test was applied with group membership as the independent variable and TTR as the dependent variable. For MATTR, a Mann-Whitney U test was applied with group membership as the independent variable and MATTR as the dependent variable. This test was used because of the non-parametric nature of TTR50, TTR, and MATTR.

2.4.b Standardized tests

Standardized test scores were collected from original researchers' data records. Since studies used different standardized tests, scores were grouped by receptive vocabulary (*PPVT* and *Receptive One-Word Picture Vocabulary Test*, picture vocabulary subtest of the *TOLD P-3*) and expressive vocabulary (*EVT*, and *Expressive One-Word Picture Vocabulary Test*, oral vocabulary subtest of the *TOLD P-3*). All scores were converted into standard scores and z-scores. Expressive and receptive vocabulary standard scores were compared between the two groups using an independent samples *t* test with group membership as the independent variable and standard scores as the dependent variable. Additionally, expressive and receptive vocabulary z scores were

compared between the two groups using an independent samples *t* test with group membership as the independent variable and z-scores as the dependent variable.

Dissociations between children's profiles on standardized vocabulary measures were determined by comparing differences of standard scores between receptive and expressive. A difference of more than 1 standard deviation in standard score between a child's expressive and receptive scores was considered a dissociation. The number of dissociations between the CWS and the CWNS was compared using a Pearson's chi-square test.

2.4.c Relationships among lexical diversity values in spontaneous speech and standardized test scores

The correlations among the spontaneous lexical diversity measures (NDW, TTR, VocD, MATTR) and standardized measures (expressive and receptive vocabulary measures) were computed to determine which language sample measure(s) related most strongly to standardized test scores. A Spearman correlation matrix was applied with the spontaneous language measures as the independent variable and expressive vocabulary standard scores of both groups combined as the dependent variable.

3. Results

3.1 Lexical diversity in spontaneous speech of CWS and CWNS

For this set of comparisons we set alpha at $p = 0.005$ after correcting for multiple comparisons. In this case, Bonferroni adjustment divided alpha = .05 by 9 to obtain an adjusted value of 0.005. An independent samples *t* test was applied to the two groups' NDW and VocD values. A Mann-Whitney U test was applied to the two groups' TTR, TTR50, and MATTR, due to lack of homogeneity of variance. Contrary to hypotheses,

the mean number of different words (NDW) for CWS (50.89) and CWNS (50.66) were remarkably similar ($t(148) = 0.19, p = 0.85$). Similarly, VocD in samples from the CWS (64.23) and the CWNS (66.46) did not differ ($t(150) = -0.66, p = 0.51$). Type-token ratio (TTR) was also not significantly different between the CWS (0.257) and CWNS (0.271) ($Z = -1.85, p = 0.06$). The type-token ratio of the middle 50 utterances (TTR50) was also not significant between the CWS (0.476) and CWNS (0.488) ($Z = -1.42, p = 0.16$). Finally, the MATTR between the groups was very similar between the CWS (0.671) and CWNS (0.677) ($Z = -0.43, p = 0.67$).

The children who stutter had a greater number of utterances, but this was not significant ($t(150) = 2.51, p = 0.013$). The CWS had an average of 217 utterances while the CWNS had an average of 176 utterances. CWS also had more tokens, but this was also not significant ($t(150) = 1.71, p = 0.09$). The CWS had an average of 847 tokens and the CWNS had an average of 713 tokens. Despite the CWS having a greater number of utterances and tokens, they displayed a remarkably similar MLU (4.165) as CWNS (4.166).

Table 2: Mean lexical diversity for CWS and CWNS across measures

	TTR	TTR50	NDW	VocD	MATTR
CWS	0.26 (SD = 0.09)	0.48 (SD = 0.07)	50.89 (SD = 7.83)	64.23 (SD = 19.10)	0.67 (SD = 0.06)
CWNS	0.27 (SD = 0.08)	0.49 (SD = 0.07)	50.66 (SD = 7.26)	66.46 (SD = 22.37)	0.68 (SD = 0.06)

3.2 Standardized tests

An independent samples *t* test with group membership as the independent variable and standard scores and z-scores as the dependent variable was applied to 50 pairs of CWS and CWNS. CWS performed significantly worse than CWNS on both receptive and expressive vocabulary tests. On expressive standard scores, the CWS achieved an average mean score of 108.16, while the CWNS had an average mean score of 117.42 ($t(98) = -3.56, p = 0.00057$). On expressive z-scores, the CWS scored at an average of 0.576 while the CWNS displayed an average of 1.21 ($t(98) = -3.68, p = 0.00039$), a difference of almost one standard deviation. On receptive standard scores, the CWS scored at an average mean of 109.18 while the CWNS achieved an average mean of 115.08 ($t(98) = -2.02, p = 0.05$). On receptive z scores, the CWS showed an average of 0.65 while the CWNS had an average of 1.038 ($t(98) = -1.99, p = 0.05$). See Figure 2 for a graphic representation of CWS and CWNS performance on standardized vocabulary tests. All participants performed above average on the standardized measures, which could be because the majority of the participants were from higher SES backgrounds (Nippold, 2012).

In a post-hoc analysis using a two-way ANOVA, we found greater differences on standardized test scores between the CWS and CWNS in the older age group (above age 4) than in the younger age group (below age 4). These differences, although not significant, were present both expressively ($F(1, 50) = 0.24, p = 0.63$) and receptively ($F(1, 50) = 1.28, p = 0.26$).

A chi-square test of independence was performed to examine the relation between stuttering and dissociations of receptive and expressive vocabulary. The relation between

these variables was not significant, $X^2(2, N = 100) = 0.053, p = 0.817$. Children who stutter were not more likely to exhibit dissociations between performance on receptive vocabulary tests and expressive vocabulary tests. The number was almost exactly equal between the groups, with 12 CWS and 13 CWNS exhibiting dissociations. Of the 12 CWS, 4 exhibited a higher expressive vocabulary score, and 8 exhibited a higher receptive vocabulary score. Of the 13 CWNS, 8 exhibited a higher expressive vocabulary score, and 5 exhibited a higher receptive vocabulary score. This means that of those that exhibited expressive-receptive vocabulary dissociations, most typically fluent children performed better on the expressive vocabulary test, while most CWS scored better on the receptive vocabulary test.

Figure 2: Standardized test performance

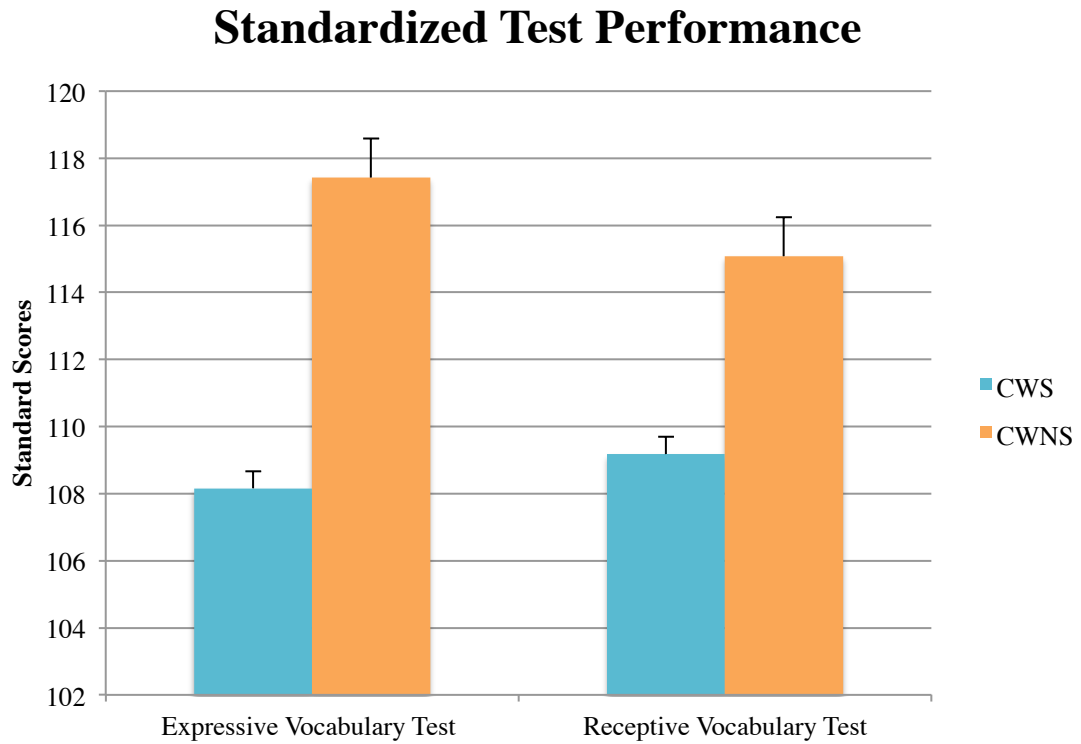


Table 3: Dissociations between receptive and expressive vocabulary

GROUP	DISSOCIATION		
	NO	YES	TOTAL
CWS	38	12	50
CWNS	37	13	50
TOTAL	75	25	100

3.3 Comparing lexical diversity in spontaneous speech vs. standardized tests

A correlation matrix was constructed with the spontaneous language measures as the independent variable and expressive vocabulary standard score as the dependent variable. For this set of comparisons we set alpha at $p = 0.005$ after correcting for multiple comparisons. In this case, Bonferroni adjustment divided alpha = .05 by 10 to obtain an adjusted value of 0.005. Among lexical diversity measures, there was a correlation between TTR and TTR50 ($r(52) = 0.37, p = 0.0067$), VocD and NDW ($r(52) = 0.45, p = 0.0008$), VocD and MATTR ($r(52) = 0.72, p < 0.0000001$), and NDW and MATTR ($r(52) = 0.62, p = 0.000001$). There was a correlation between MATTR and TTR50, but after Bonferroni correction, this was not significant ($r(52) = 0.29, p = 0.03$). The only lexical diversity measure that correlated with expressive standard scores on standardized tests was TTR50, but this was not significant after Bonferroni correction ($r(52) = 0.34, p = 0.01$). There was a slight correlation between MATTR and expressive standard scores, but this was also not significant ($r(52) = 0.23, p = 0.09$). See figures 3-6 for graphic representations of these correlations.

Table 4: Correlations among lexical diversity measures

	Overall TTR	NDW100	VocD	MATTR
NDW100	$r = 0.13$ $p = 0.34$			
VocD	$r = -0.09$ $p = 0.51$	$r = 0.45^*$ $p = 0.00008$		
MATTR	$r = 0.25$ $p = 0.007$	$r = 0.62^*$ $p = 0.000001$	$r = 0.72^*$ $p = <0.00001$	
TTR50	$r = 0.37$ $p = 0.007$	$r = -0.01$ $p = 0.93$	$r = 0.13$ $p = 0.37$	$r = 0.29$ $p = 0.03$

* = significant at 0.005

Figure 3: Correlation between NDW and VocD

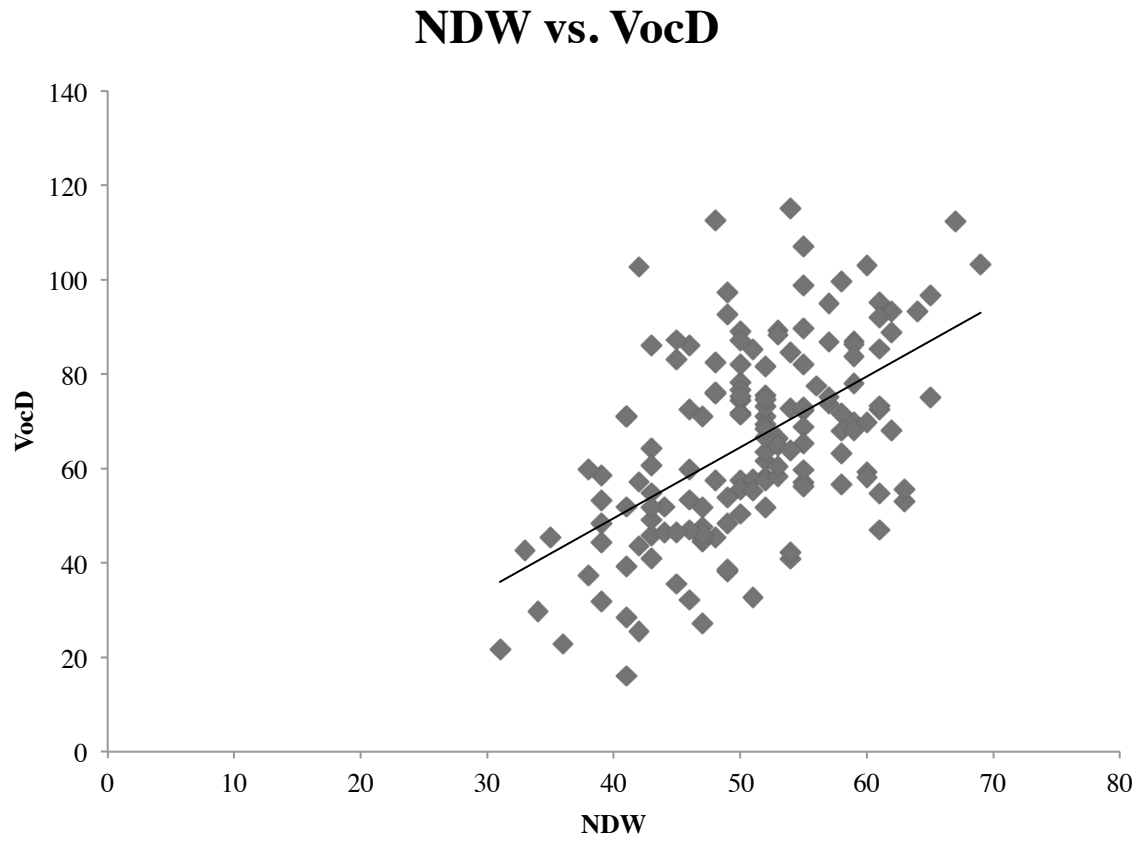


Figure 4: Correlation between NDW and MATTR

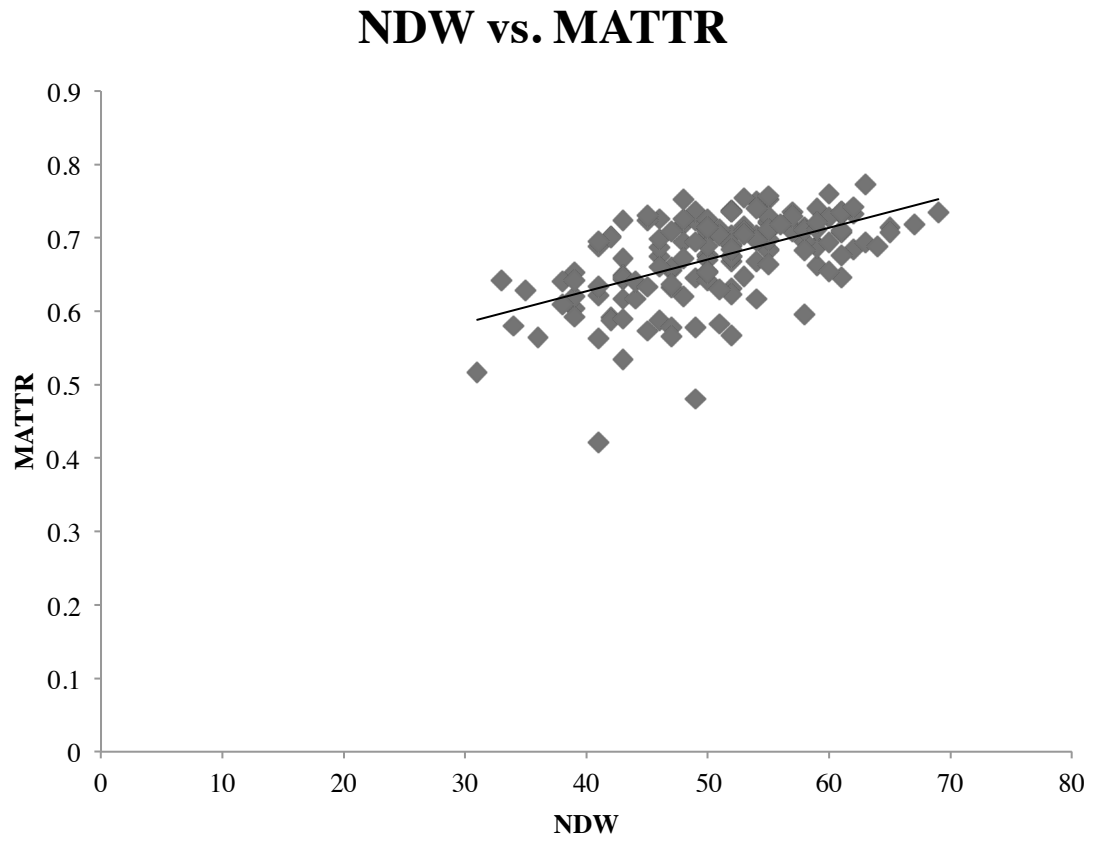


Figure 5: Correlation between VocD and MATTR

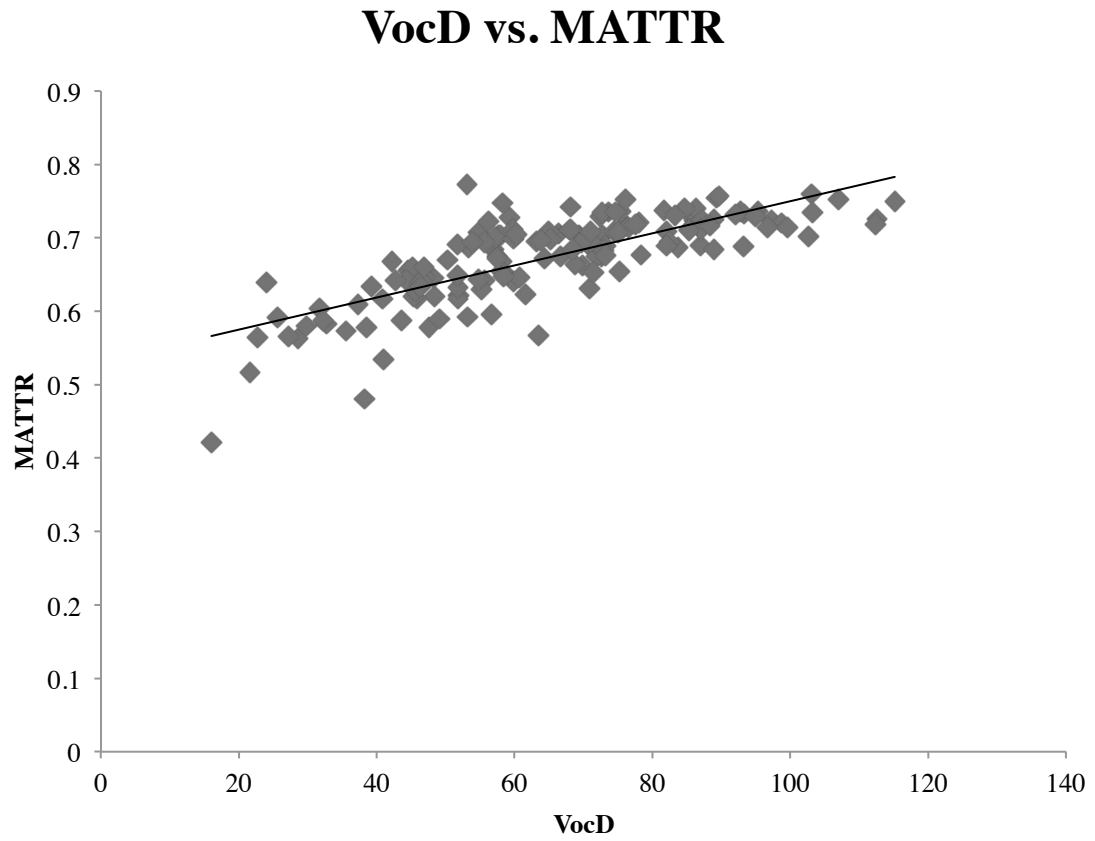
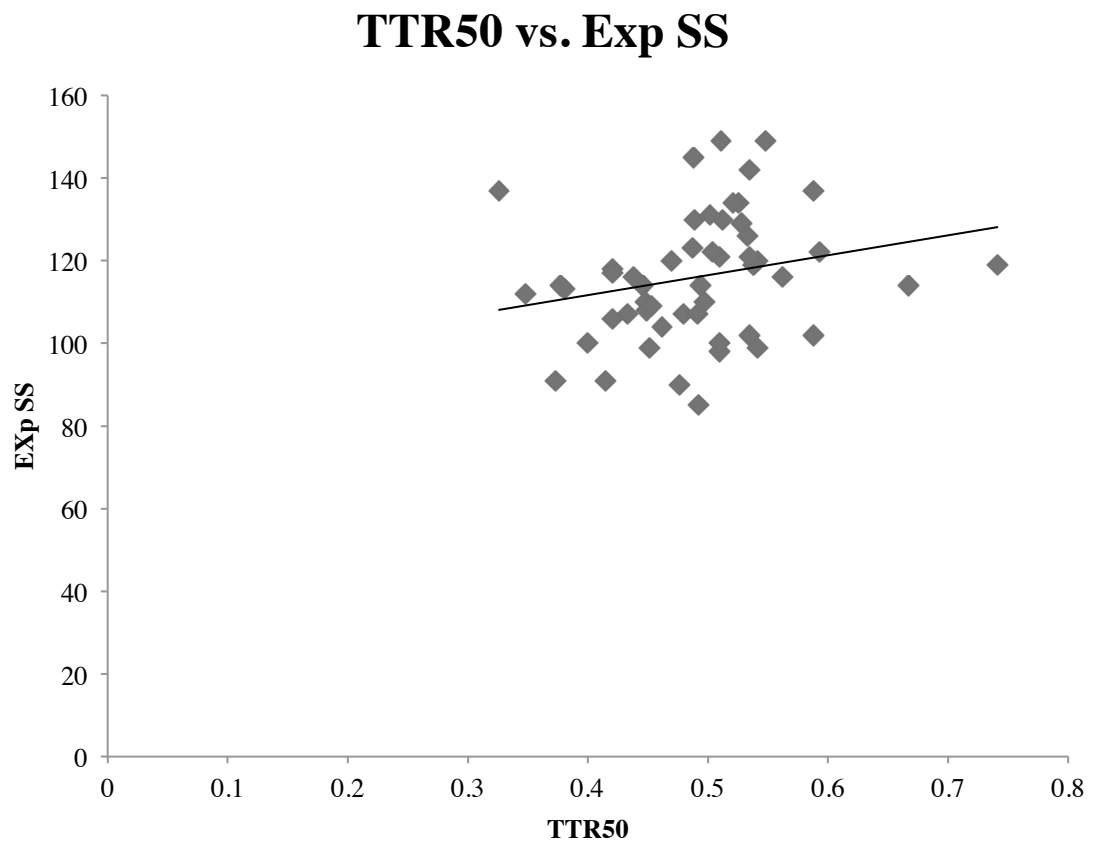


Table 5: Correlations between lexical diversity measures and expressive vocabulary

		Overall TTR	NDW	MATTR	TTR50	VocD
Expressive Vocabulary SS	Correlation (r)	0.004	0.001	0.23	0.34	0.12
	p-value	0.98	0.99	0.09	0.01	0.40

Figure 6: Correlation between expressive vocabulary standardized test score and TTR50



4. Conclusions

The primary goal of this study was to compare lexical skills of children who stutter and matched peers in both spontaneous speech and on standardized vocabulary tests. This study used a large sample of matched pairs from several different studies. Dissociations were examined in order to determine if children who stutter have imbalanced language skills consistent with the Demands and Capacities Model (Adams, 1990; Starkweather & Gottwald, 1990; Starkweather, 1987). Because of the wide range of lexical diversity measures used across studies, a secondary goal was to assess the validity of measures of lexical diversity in spontaneous speech. It was hypothesized that the different measures used by researchers could be contributing to the conflicting results across studies.

4.1 Lexical diversity in spontaneous speech

Contrary to hypotheses, children who stutter demonstrated quite similar lexical diversity in spontaneous speech when compared to children who do not stutter. This is consistent with some previous literature, which found no differences on NDW or TTR between CWS and CWNS (Watkins & Yairi, 1997; Watkins et al., 1999; Bonelli, 2001; Silverman & Bernstein Ratner, 2002). While Silverman and Bernstein Ratner (2002) found significant differences in VocD between the groups, this study did not. This is surprising, considering analysis of VocD was replicated identically to that study (on the entire sample), and subjects from that study were among the children used in this analysis. This study employed a very large sample (N = 152), of which the Silverman and Bernstein Ratner (2002) corpus constituted only about 20% of observations (N = 30). In sum, CWS did not differ from fluent peers as assessed on five different measures of

lexical diversity, a finding suggesting that children who stutter have equivalent expressive language skill, as measured by lexical diversity, in spontaneous speech.

The CWS tended to produce more utterances, but this difference was not significant. Despite the greater number of utterances, MLU was very similar across the two groups, with only a difference of 0.001 between the two means. This is consistent with a recent study, Watts et al. (2017), who also employed a large sample of 181 children who stutter. Whereas Watts et al. (2017) compared the children who stutter to a normative database, this study individually matched children who stutter with peers identical in age, gender, and SES, allowing conclusions to be drawn with more certainty.

It is possible that sample size, as measured by number of utterances, is also an artifact of the data collection process in the various studies. CWS, as opposed to typically-developing children, may have engaged in longer adult-child interactions to provide greater opportunities to observe stuttering behaviors. This is merely speculation, but we note that sample size in utterances is not, by itself, either a research or clinical measure, beyond examination of minimally verbal children.

4.2 Lexical diversity on standardized tests

A highly significant difference was found between CWS and CWNS on expressive and receptive vocabulary scores. The CWS performed on average almost one standard deviation more poorly than the CWNS on the expressive vocabulary tests they were given across the studies in which they participated. This finding is interesting because three different vocabulary tests were used across studies, but CWS demonstrated consistent expressive and receptive vocabulary weaknesses on such tasks. Because we employed a varied age group, it appears that these differences are present from close to

stuttering onset until age eight. Post-hoc analyses displayed greater differences between the two groups in the older age group (above age 4) than in the younger age group (below age 4). This is expected, given that the younger group by definition contains more children likely to recover from stuttering, who should more closely resemble typical children in most respects.

Only two studies to date have found poorer performance of CWS on **both** expressive and receptive vocabulary as measured by standardized tests (Coulter et al., 2009; Choo et al., 2016). Similar to this study, these studies both employed a large number of participants (85 pairs, Coulter et al.) (66 CWS and 53 CWNS, Choo et al.). It should be cautioned that 20 pairs from Choo et al. were used in this analysis, making up almost 50% of the sample, so it is not surprising we found similar results. Coulter et al., (2009), however, employed 45 pairs from Anderson et al., (2005); although those authors found significant differences on expressive vocabulary, they did not find differences on receptive vocabulary (as measured by the *PPVT*). The fact that Coulter et al., (2009) found differences by adding 40 more pairs to the analysis attests to the power and importance of having a large sample size in making judgments about language skills in CWS.

Consistent with previous research, both CWS and CWNS groups still fell within the normal range for average test scores, a finding that supports the notion that most children who stutter are not language-impaired, per se, but instead display fairly subtle language weaknesses. In fact, only two CWS performed less than one standard deviation below the mean on any analyses, and these scores were on a receptive vocabulary test. Relevant to past characterizations of “dissociations”, both of these children had a

standard score of above 100 on the expressive vocabulary test, exhibiting an expressive-receptive gap (where expressive skills are greater than receptive skills).

At first glance, it seems rather surprising that we should find differences in lexical skills between the groups on standardized tests but not in spontaneous language. It appears that there may be methodological differences between the two tasks that influence performance. First, language testing does not permit self-selection of lexicon. Children may not have been exposed to the words that are on the test, or find them difficult to retrieve or resolve among comprehension options. It is unclear whether this breakdown might occur while retrieving the concept, phonological representation, or motor sequencing pattern of the word. Spontaneous play gives children free range in the words that they use. There are no set lexical targets required to satisfy the task.

Second, language testing demands attention, motivation, and good processing skills. The children who stutter used in these studies could have had weaker executive functioning skills, impairing language performance on standardized tests, such as attention, or short-term or phonological working memory. Such deficits have been observed in other studies (Reilly & Donaher, 2005; Hakim & Bernstein Ratner, 2004; Pelczarski & Yaruss, 2016; Anderson & Wagovich, 2010). For instance, skills such as memory for specific vocabulary entries could contribute to inferior performance on a closed set of items, whereas in a language sample, children are free to choose any appropriate easily retrievable words. In fact, research shows that CWS perform significantly worse on nonword repetition tasks (Anderson & Wagovich, 2010; Hakim & Bernstein Ratner, 2004) and phonological memory tasks (Pelczarski & Yaruss, 2016), both of which correlate with performance on standardized language measures.

Children who stutter were no more likely to exhibit dissociations than children who do not stutter. This is inconsistent with previous research demonstrating that CWS exhibit more frequent linguistic dissociations than CWNS (Anderson and Conture, 2000; Anderson et al., 2005; Coulter, 2009). Of those who did exhibit dissociations, the typically fluent peers tended to perform better on the expressive vocabulary measures, while the CWS did better on the receptive vocabulary test. This result should be viewed with caution, however, as the sample size of children showing marked discrepancies between receptive and expressive language skills was very small. This is an opposite pattern found by Coulter et al. (2009), where they found that CWS generally did better on the expressive language tasks. However, not only did Coulter et al. (2009) identify more participants who exhibited greater dissociations (allowing for a greater sample size), but they also tested and defined dissociations for overall language abilities, whereas we were testing and classifying only on lexical tasks. Moreover, it should also be noted that the pairs in Coulter's study were matched by mean standard scores on the *PPVT* and *EVT* in order to control whether differences in vocabulary measures could influence dissociations. In essence, Coulter et al. blocked potential vocabulary effects and very few conclusions can be made about the vocabulary skills of the CWS in Coulter et al.'s study. It could be that greater dissociations are evident on tasks when CWS are asked to process vocabulary within larger, syntactic environments.

4.3 Comparing lexical diversity in spontaneous speech vs. standardized tests

Among spontaneous lexical diversity measures, we found high correlations among TTR and TTR50, VocD and NDW, VocD and MATTR, and NDW and MATTR. Therefore, NDW, VocD, and MATTR appear to correlate with each other across the

board for children in the age range we studied, leaving TTR and TTR50 as the least comparable when compared to other measures. This is interesting, as TTR is the oldest and most commonly used measure among clinicians when measuring lexical diversity in spontaneous speech (Bernstein Ratner & MacWhinney, 2016). Even when sample size is standardized, TTR still appears to paint a different picture of individual lexical skills than the other three measures, even though all measures were consistent in regards to group differences. The correlations between VocD and MATTR are not surprising, given that both have proved more reliable across sample size than other measures (Fergadiotis et al., 2013; Fergadiotis et al., 2015; Covington & McFall, 2010; Silverman & Bernstein Ratner, 2002; Malvern et al., 2004; McCarthy & Jarvis, 2007; Koizumi & In'nami, 2012) and both use the entire sample in their analysis. The high correlation of NDW with VocD and MATTR, however, is surprising, given that NDW only used the first 100 words of a sample. It could be that using the beginning of a transcript produces less variable results than using the middle of a transcript where the start and end points are arbitrarily chosen for analysis. Also, since NDW can be found to differ depending on which part of the transcript is used for analysis (Malvern et al., 2004), there is a possibility that we would get a different result if we had calculated NDW in a different part of the transcript.

The only measure that correlated strongly with scores on standardized expressive vocabulary measures was TTR50. It is not surprising that overall TTR was not correlated with other measures or concurrently with scores on standardized expressive vocabulary measures (Hess et al., 1984, Hess et al., 1989; Richards, 1987; Malvern et al., 2004; Koizumi & In'nami, 2012; Bernstein Ratner & MacWhinney, 2016), but it is surprising that we found TTR50 to be correlated with scores on expressive vocabulary measures.

This is because previous research suggests even when TTR is standardized, it is not correlated with scores on vocabulary measures (Silverman & Bernstein Ratner, 2002). The correlation between TTR50 and results on standardized expressive vocabulary measures is also surprising given that expressive standard scores displayed a difference between the two groups, whereas TTR50 did not. However, this finding is similar to that seen in the literature on specific language impairment (SLI), where a difference has been found between children with SLI and controls on standardized measures but not on TTR of a fixed length (in that study, 100 utterances rather than 50, which is the standard definition of NDW) (Watkins et al., 1995).

This leaves a puzzling question – why is TTR50 most highly correlated with standardized expressive vocabulary scores, but shows the least strong correlation with other language sample analysis measures? Although a correlation was present, it was still a modest correlation at $r = 0.34$ and was not significant after correcting for multiple comparisons. By comparison, MATTR and VocD both showed a positive correlation with expressive standard scores. NDW and overall TTR both showed close to no correlation with language test scores ($r = 0.004$ and 0.001 , respectively). The observed correlation between TTR50 and expressive standard scores could be due to our methodology in choosing the middle 50 utterances from the child language samples, since we know that TTR varies depending on which part of the sample is used (Malvern et al., 2004; Silverman & Bernstein Ratner, 2002; McCarthy, 2007). All children were given a standardized set of toys to play with, which we felt made truncating the sample in random places risky. Those truncated samples may reveal a skewed lexical diversity depending on the activity the child was engaged in while the sample was being recorded.

As pure speculation, we might hypothesize that TTR50 measures diversity in language that accompanies the child's play with the first and most attractive play options available at the time.

4.4 Limitations

The potential limitations of this study include the wide age range of participants and the variability in conducting language samples across studies. In regards to the first issue, although efforts were made to examine children who stutter as close to onset as possible, the diversity of studies instead allowed us to pool a wide range of participants from close to onset through age eight. This makes it difficult to rule out whether the observed standardized test differences were the cause or result of stuttering. However, finding an effect amongst such a large sample across ages increases power and generalizability. More importantly, because there was no difference observed in lexical diversity of spontaneous language and no significant effect of age on the language differences between CWS and CWNS, it is unlikely that word avoidances and substitutions conditioned by stuttering should have impacted lexical diversity.

Pertinent to the second issue, some studies we included analyzed mother-child interactions, while others used clinician-child interactions for their language samples. To help remediate this, pairs were matched in this study as much as possible to match for parent- or clinician-led language sample context. Because this study pooled data from different studies, clinicians could have differed in what they asked and did with the child. Even though the toys and topics may have differed, it is expected that because the pairs were matched on several factors and the sample was very large, these differences should have evened out and not created large impacts on our findings.

4.5 Future research

Future studies could divide participants into children near onset and older children to help determine if lexical skills change over time since onset. Even more informative would be longitudinal studies, in which factors such as persistence and recovery can also be determined from potential linguistic measures, as are now suggested by growth-modeling over time of a subset of the current study population, as reported by Leech et al., in press.

Since TTR50 was found to be most highly correlated with expressive vocabulary standardized scores, but NDW was found to be best inter-correlated with other measures, future research could compare NDW and TTR when computed on different parts of the language sample (e.g., first 100 words, middle 50 utterances, last 100 utterances, etc.) to see if this leads to different results. Certainly, TTR50 would require a shorter language sample to compute (although grammatical analysis procedures usually performed on the same language sample, such as IPSYN, still require a minimal sample of 100 utterances).

To help understand why we observed differences on standardized measures and not spontaneous measures, future research can examine executive functioning skills in CWS including memory and attention, and the effects that they have on standardized test performance. In order to do this, studies would need to gather both extensive language measures and concurrent executive function indicators, something not done for the majority of children examined in this analysis. Also, because it is unclear what exactly is difficult for children who stutter to retrieve, future studies could do a more sophisticated assessment of lexical retrieval during standardized measures. This may involve performing an item-by-item analysis on standardized measures in order to determine if

the CWS are hitting the ceiling earlier than CWNS and therefore having a lower number of opportunities or if they are making a lot of errors with a similar number of opportunities.

4.6 Summary

The primary focus of this study was to examine lexical diversity in spontaneous conversational speech in language samples as well as standardized language testing, by collapsing six longitudinal studies of children who stutter. The secondary focus was to assess the comparability of different lexical diversity measures in assessment of children's lexical profiles in conversational speech.

Children who stutter demonstrated remarkably similar lexical diversity compared to children who do not stutter in spontaneous speech on several LSA measures: overall TTR, TTR50, NDW, MATTR, and VocD. The CWS tended to have had more utterances in their research language samples, but this difference was not significant and may have been due to original efforts to collect more data from CWS for fluency analyses. Despite the greater number of utterances, length of average utterances, as measured by MLU, was very similar across the two groups.

A highly significant difference was found between CWS and CWNS on expressive and receptive vocabulary scores across studies that used different vocabulary measures. The CWS performed, on average, almost one standard deviation below the CWNS on the expressive vocabulary tests. Both groups still fell within normal range, a finding that supports that children who stutter are not, as a group, frankly language-impaired. Children who stutter were found to be no more likely to exhibit dissociations on expressive and receptive vocabulary than children who do not stutter.

Among spontaneous lexical diversity measures, there were significant correlations among NDW, VocD, and MATTR, suggesting the value of using the entire transcript for language analysis. However, the only measure that correlated with scores on standardized expressive vocabulary for these children was TTR50, a surprising finding given that TTR50 is not correlated with any of the other measures, and uses a much smaller proportion of the available language sample for analysis.

Taken together, our research provides possible future directions for research in stuttering. These could include analysis that goes beyond lexical profiles and examines grammatical or syntactic profiles. Our work also suggests possible value in future work to investigate best measures of lexical productivity in language sample analysis of children who both do and do not stutter.

References

- Adams, M. R. (1990). The demands and capacities model I: theoretical elaborations. *Journal of Fluency Disorders, 15*, 135–141.
- Ammons, R. B. & Ammons, H. S. (1948). *The full range picture vocabulary test*. New Orleans: R. B. Ammons.
- Anderson, J. D. & Conture, E. G. (2000). Language abilities of children who stutter: A preliminary study. *Journal of Fluency Disorders, 25*, 283-304.
- Anderson, J. D., Pellowski, M. W., & Conture, E. G. (2005). Childhood stuttering and dissociations across linguistic domains. *Journal of Fluency Disorders, 30*, 219-253.
- Anderson, J. D. & Wagovich, S. A. (2010). Relationships among linguistic processing speed, phonological working memory, and attention in children who stutter. *Journal of Fluency Disorders, 35*, 216-234.
- Bates, E., Appelbaum, M., Salcedo, J., Saygin, A. P., & Pizzamiglio, L. (2003). Quantifying dissociations in neuropsychological research. *Journal of Clinical and Experimental Neuropsychology, 25*, 1128–1153.
- Bernstein, N. E. (1981). Are there constraints on childhood disfluency? *Journal of Fluency Disorders, 6*, 341-350.
- Bernstein Ratner, N. & Costa Sih, C. (1987). Effects of gradual increases in sentence length and complexity on children's dysfluency. *Journal of Speech and Hearing Disorders, 52*, 278-287.

- Bernstein Ratner, N. & MacWhinney, B. (2016). Your laptop to the rescue: Using the child language data exchange system archive and CLAN utilities to improve language sample analysis. *Seminars in Speech and Language, 37*, 74-84.
- Bloodstein, O. & Grossman, M. (1981). Early stutterings: some aspects of their form and distribution. *Journal of Speech and Hearing Research, 24*, 298-302.
- Bonelli, P., Dixon, M., & Bernstein Ratner, N. (2000). Child and parent speech and language following the Lidcombe Programme of early stuttering intervention. *Clinical Linguistics & Phonetics, 14*, 427-446.
- Brownell, R. (2000). *Receptive one-word picture vocabulary test*. Novato, CA: Academic Therapy Publications.
- Carroll, J. B. (1964). *Language and thought*. Englewood Cliffs, NJ: Prentice-Hall.
- Choo, A. L., Burnham, E., Hicks, K., & Chang, S. (2016). Dissociations among linguistic, cognitive, and auditory-motor neuroanatomical domains in children who stutter. *Journal of Communication Disorders, 61*, 29-47.
- Coulter, C. E., Anderson, J. D., & Conture, E. G. (2009). Childhood stuttering and dissociations across linguistic domains: a replication and extension. *Journal of Fluency Disorders, 34*, 257-278.
- Covington, M. A. & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics, 17*, 94-100.
- Crystal, D. (1987). Towards a 'bucket' theory of language disability: taking account of interaction between linguistic levels. *Clinical Linguistics & Phonetics, 1*, 7-22.

- Dale, P., Bates, E., Reznick, S., & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language, 16*, 239-249.
- Dawson, J. & Stout, C. (2003). *The structured photographic expressive language test* (3rd ed.). DeKalb, IL: Janelle Publications.
- Dickinson, D. K. & Tabors, P. O. (2001). *Beginning literacy with language: Young children learning at home and school*. Baltimore: Paul Brookes Publishing.
- Dugast, D. (1978). Sur quoi se fonde la notion d'étendue théorique du vocabulaire? *Le Français Moderne, 46*, 25–32.
- Dugast, D. (1979). *Vocabulaire et stylistique. I Théâtre et dialogue. Travaux de linguistique quantitative*. Geneva: Slatkine-Champion.
- Dunn, L.M. & Dunn, D. M. (1997). *Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service Inc.
- Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service, Inc.
- Dunn, L. M. & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Minneapolis, MN: American Guidance Service Inc.
- Ellis Weismer, S., Venker, C., Evans, J. L., & Moyle, M. (2013). Fast mapping in late-talking toddlers. *Applied Psycholinguistics, 34*, 69-89.
- Fenson, L., Dale, P. S., Reznick, J. S. (1993). *The Macarthur communicative development inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.

- Fergadiotis, G. (2011). Modeling lexical diversity across language sampling and estimation techniques. (Dissertation). Retrieved from Arizona State University Digital Repository.
- 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*, 397–408.
- Fergadiotis, G., Wright, H. H., & Green, S. B. (2015). Psychometric evaluation of lexical diversity indices: assessing length effects. *Journal of Speech, Language, and Hearing Research, 58*, 840-852.
- 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*, 37-42.
- Gardner, M. (2000). *Expressive one-word picture vocabulary test, 2000 edition*. Austin, TX: Pro-Ed.
- Goldman, R. & Fristoe, M. (1986). *Goldman-fristoe test of articulation (GFTA)*. Circle Pines, MN: American Guidance Service Inc.
- Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique*. Dordrecht: D. Reidel
- Hakim, H. B. & Bernstein Ratner, N. (2004). Nonword repetition abilities of children who stutter: an exploratory study. *Journal of Fluency Disorders, 29*, 179-199.
- Hall, N. E., Higgins, K., Wagovich, S. A., Farkas, L., Cote, E., Russell, L., & Ward, J. (2007). A developmental study of normal fluency and language. In J. Au-Yeung

- & M.M. Leahy (Eds.) *Research, treatment, and self-help in fluency disorders: New horizons*. (pp. 130-135). Dublin, Ireland: International Fluency Association.
- Hall, N. E. (2004). Lexical development and retrieval in treating children who stutter. *Language, Speech, and Hearing Services in Schools, 35*, 57-69.
- Hall, N. E., Wagovich, S. A., & Bernstein Ratner, N. (2007). Language considerations in developmental stuttering. In E. Conture, & R. Curlee (Eds.), *Stuttering and related disorders of fluency* (3rd ed., pp. 153–167). New York: Thieme.
- Hess, C. W., Ritchie, K. P., & Landry, R. G. (1984). The type-token ratio and vocabulary performance. *Psychological Reports, 55*, 51-57.
- Hess, C. W., Haug, H. T., & Landry, R. G. (1989). The reliability of type-token ratios for the oral language of school age children. *Journal of Speech and Hearing Research, 32*, 536-540.
- Hollister, J., Alpermann, A., & Zebrowski, P. (2012). Dissociations within and across linguistic motor domains in children who stutter. *Paper presented at the American Speech-Language Hearing Association*. Atlanta, GA.
- Hresko, W., Reid, D., & Hamill, D. (1991). *Test of Early Language Development-2 (TELD-2)*. Austin, Tx: PRO-ED.
- Johnson, W. (1944). Studies in language behavior: I. A program of research. *Psychological Monographs, 56*, 1-15.
- Johnson, K.N., Karrass, J., Conture, E.G., & Walden, T. (2009) Influence of stuttering variation on talker group classification in preschool children: Preliminary findings. *Journal of Communication Disorders, 42*, 195–210.

- Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders, 12*, 28-41.
- Koizumi, R. & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System, 40*, 554-564.
- Leadholm, B. & Miller, J. F. (1992). *Language sample analysis: The Wisconsin guide*. Madison, WI: Wisconsin Department of Public Instruction.
- Leech, K. A., Bernstein Ratner, N., Brown, B., & Weber, C. M. (in press). Preliminary evidence that growth in productive language differentiates childhood stuttering persistence and recovery. *Journal of Speech, Language, and Hearing Research*.
- Logan, K. J. & Conture, E. G. (1995). Length, grammatical complexity, and rate differences in stuttered and fluent conversational utterances of children who stutter. *Journal of Fluency Disorders, 20*, 35-61.
- Logan, K. J. & Conture, E. G. (1997). Selected temporal, grammatical and phonological characteristics of conversational utterances produced by CWS. *Journal of Speech, Language, and Hearing Research, 40*, 107-120.
- MacPherson, M. K. & 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*, 89-102.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Malvern, D. D., Richards, B. J., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development*. New York: Palgrave MacMillan.

- 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). Clevedon: Multilingual Matters.
- McCarthy, P. M. & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing*, 24, 459-488.
- McCarthy, P. M. & Jarvis, S. (2010). MTLTD, Voc-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381-392.
- McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literacy and Linguistic Computing*, 15, 323-337.
- Millager, R. A., Conture, E. G., Walden, T. A., & Kelly, E. M. (2014). Expressive language intratest scatter of preschool-age children who stutter. *Contemporary Issues in Communication Sciences and Disorders*, 41, 110-119.
- Miller, J. F. (1991). Quantifying productive language disorders. In J. F. Miller (Ed.), *Research on child language disorders: A decade of progress*. Austin, Texas: Pro-Ed.
- Miller, J. F. (1981). *Assessing language production: experimental procedures*. London: Arnold.
- Newcomer, P. & Hammill, D. (1982). *Test of language development-primary*. Austin, Texas: Pro-Ed.
- Newcomer, P. & Hamill, D. (1997). *Test of language development*. Austin, Texas: Pro-Ed.

- Nippold, M. A. (1990). Concomitant speech and language disorders in stuttering children: A critique of the literature. *Journal of Speech and Hearing Disorders*, 55, 51-60.
- Nippold, M. A. (2012). Stuttering and language ability in children: questioning the connection. *American Journal of Speech-Language Pathology*, 21, 183-196.
- 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, 163-179.
- Owen, A. J. & Leonard, L. B. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: application of D. *Journal of Speech, Language, and Hearing Research*, 45, 927-937.
- Pelczarski, K. M. & Yaruss, J. S. (2016). Phonological memory in young children who stutter. *Journal of Communication Disorders*, 62, 54-66.
- Perkins, W. H., Kent, R. D., & Curlee, R. F. (1991). A theory of neuropsycholinguistic function in stuttering. *Journal of Speech and Hearing Research*, 34, 734-752.
- PerozziI, J. A. & Kunze, L. H. (1969). Language abilities of stuttering children. *Folia Phoniatica*, 21, 386-392.
- Retherford, K. S. (2000). *Guide to analysis of language transcripts (3rd ed.)*. Eau Claire, WI: Thinking Publications.
- Richards, B. (1987). Type/token ratios: what do they really tell us? *Journal of Child Language*, 14, 201-209.

- Richels, C., Buhr, A., Conture, E., & Ntourou, K. (2010). Utterance complexity and stuttering on function words in preschool-age children who stutter. *Journal of fluency disorders, 35*, 314-331.
- Ryan, B. P. (1992). Articulation, language, rate, and fluency characteristics of stuttering and nonstuttering preschool children. *Journal of Speech and Hearing Research, 35*, 333–342.
- Sawyer, J., Chon, H., & Ambrose, N. G. (2008). Influences of rate, length, and complexity on speech disfluency in a single-speech sample in preschool children who stutter. *Journal of fluency disorders, 33*, 220-240.
- Silverman, S. & Bernstein, N. (2002). Measuring lexical diversity in children who stutter: application of vocd. *Journal of Fluency Disorders, 27*, 289-304.
- Smith, A., Goffman, L., Sasisekaran, J. & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: evidence from behavioral and kinematic indices of nonword repetition performance. *Journal of Fluency Disorders, 37*, 344–358.
- Snow, C. E. (1989). Imitativeness: a trait or a skill? In G. E. Speidel and K. E. Nelson (Eds), *The many faces of imitation in language learning* (pp. 73-90). New York: Springer-Verlag.
- Somers, H. H. (1966). Statistical methods in literary analysis. In Leeds, J., editor, *The computer and literary style*. Kent, OH: Kent State University, 128–40.
- Starkweather, C. W. (1987). *Fluency and stuttering*. Englewood Cliffs, NJ: Prentice-Hall.
- Starkweather, C. W. & Gottwald, S. R. (1990). The demands and capacities model II: Clinical applications. *Journal of Fluency Disorders, 15*, 143–157.

- Templin M. C. (1957). *Certain language skills in children*. Minneapolis: University of Minnesota Press.
- Tommerdahl, J. & Kilpatrick, C. (2013). Analysing frequency and temporal reliability of children's morphosyntactic production in spontaneous language samples of varying lengths. *Child Language Teaching and Therapy*, 29, 171-183.
- Tommerdahl, J. & Kilpatrick, C. D. (2014). The reliability of morphological analyses in language samples. *Language Testing*, 31, 3-18.
- Wagovich, S. & Hall, N. (2007). Lexical and fluency changes in young children who stutter. In J. Au-Yeung & M. M. Leahy (Eds), *Research, treatment, and self-help in fluency disorders: New horizons* (pp. 141-147). Dublin, Ireland: International Fluency Association.
- Warren-Leubecker, A. & Bohannon, J. N. (1984). Intonation patterns in child-directed speech: Mother-father speech. *Child Development*, 55, 1379-1385.
- Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research*, 38, 1349-1355.
- Watkins, R. V. & Yairi, E. (1997). Language production abilities of children whose stuttering persisted or recovered. *Journal of Speech, Language, and Hearing Research*, 40, 385-399.
- Watkins, R. V., Yairi, E., & Ambrose, N. G. (1999). Early childhood stuttering III: Initial status of expressive language abilities. *Journal of Speech, Language, and Hearing Research*, 42, 1125-1135.

- Watkins, R. V. & Johnson, B. W. (2004). Language abilities in children who stutter: toward improved research and clinical applications. *Language, Speech, and Hearing Services in Schools, 35*, 82-89.
- 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*, 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.
- Westby, C. E. (1979). Language performance of stuttering and nonstuttering children. *Journal of Communication Disorders, 12*, 133-145.
- Williams, K. T. (1997). *Expressive vocabulary test (EVT)*. Circle Pines, MN: American Guidance Service Inc.
- Williams, K. T. (2007). *EVT-2: Expressive Vocabulary Test*. Pearson Assessments.
- Williams, D. E., Melrose, B. M., & Woods, C. L. (1969). The relationship between stuttering and academic achievement in children. *Journal of Communication Disorders, 2*, 87-98.
- Zackheim, C. T. & Conture, E. G. (2003). Childhood stuttering and speech disfluencies in relation to children's mean length of utterance: A preliminary study. *Journal of Fluency Disorders, 28*, 115-142.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool language scale* (4th ed.). San Antonio, Texas: Pearson Assessments.