ELSEVIER

Journal of Fluency Disorders
27 (2002) 289–304

Journal of
FLUENCY
DISORDERS

Measuring lexical diversity in children who stutter: application of *vocd*

Stacy Silverman[a,*], Nan Bernstein Ratner[b]

[a] *Department of Communication Science and Disorders, University of Missouri-Columbia, 303 Lewis Hall, Columbia, MO 65211-3873, USA*
[b] *Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742-8211, USA*

Received 25 June 2002; received in revised form 20 August 2002; accepted 29 August 2002

Abstract

There is growing but equivocal evidence that the language abilities of young children who stutter (CWS) may be depressed when compared to those of their fluent peers. In particular, the lexical skills of CWS have variously been reported to be weaker or stronger than comparison children in a number of recent studies. One source for such disagreement may be the measures used to compute lexical characteristics of these children's spoken conversations. In this study, we examined the concurrent validity of two measures of lexical diversity in spontaneous language samples, *Type-Token Ratio (TTR)* and the newly developed utility *vocd* (Malvern & Richards, 1997), using a standard test of expressive vocabulary as the comparison measure. Findings indicated that *vocd* values ("*D*") correlated well with standardized measures of expressive vocabulary, while *TTR* values did not. In addition, both the standardized measure and *vocd* revealed significantly poorer expressive lexical skills of CWS, whereas *TTR* analyses did not evidence this difference. Results are discussed in relation to the relative strength of *vocd* over *TTR* as a method for describing lexical characteristics of the spontaneous language samples of this population.

Educational objectives: The reader will learn about and be able to (1) identify several common measures of conversational vocabulary and the strengths and weaknesses of each, and (2) compare the performance of the young CWS in this study to their normally fluent

* Corresponding author. Tel.: +1-573-882-4278; fax: +1-573-884-8686.
E-mail address: silvermans@health.missouri.edu (S. Silverman).

PII: S0094-730X(02)00162-6

peers in terms of vocabulary performance on both formal and conversational measures of vocabulary.

Keywords: Developmental stuttering; Language; Vocabulary assessment; *Vocd*; *TTR*

1. Introduction

There is a longstanding literature that has examined whether the language abilities of children who stutter (CWS) are equivalent to those of children who do not stutter (see Nippold, 1990; Ratner, 1997, for reviews). A number of researchers have found a variety of linguistic abilities to be relatively depressed in young CWS when the children's performance is contrasted with that of their normally fluent peers (e.g., Anderson & Conture, 2000; Byrd & Cooper, 1989; Murray & Reed, 1977; Ratner & Silverman, 2000; Ryan, 1992, 2000; Westby, 1979). However, other studies have failed to document relative language delays or differences in CWS and even suggest a relative level of language precocity when compared to peer performance or assessment norms (Bonelli, Dixon, Ratner, & Onslow, 2000; Ratner & Sih, 1987; Watkins & Yairi, 1997; Watkins, Yairi, & Ambrose, 1999).

In attempting to reconcile findings, it is important to note that few areas of language ability have been appraised consistently across studies. Investigators have employed a variety of narrow standardized language assessments, such as the *Peabody Picture Vocabulary Test — Revised* (*PPVT-R*; Dunn & Dunn, 1981), or broad language batteries assessing numerous aspects of language understanding and use (cf. Anderson & Conture, 2000; Ryan, 1992, 2000; Westby, 1979); still others have employed experimental measures of performance (e.g., sentence imitation as in Gordon & Luper, 1989; Pearl & Bernthal, 1980; Ratner & Sih, 1987; Silverman & Ratner, 1997). Additional studies have examined spontaneous expressive language tendencies (Gaines, Runyan, & Meyers, 1991; Logan & Conture, 1995; McLaughlin & Cullinan, 1989; Weiss & Zebrowski, 1992). Among the indices of language ability that have been appraised in spontaneous language analyses are *Mean Length of Utterance* (*MLU*), *Developmental Sentence Score* (*DSS*), and lexical diversity, measured variously by *Type-Token Ratio* (*TTR*), *Number of Different Words* (*NDW*) and *Lexical Rarity* (*LR*; Miles & Ratner, 2001; Ratner & Silverman, 2000).

The study of lexical skills of young children who stutter is theoretically motivated by such models as Demands and Capacities (Neilson & Neilson, 1987; Starkweather, 1987), that predict that fluency breaks down when communication demands exceed individual capacities. One prediction of this model is that, if particular language skills (e.g., vocabulary skills) are weaker for an individual child compared to peers, a fluency breakdown would result, even in the face of communication demands similar to those experienced by peers. This model is particularly compelling in explaining the stuttering of young children, because the onset of the disorder co-occurs with the challenges of early language development. Multifactoral models (e.g., Smith & Kelly, 1997; Wall & Myers, 1995) also would seem to predict potential differences, because they highlight variables, including linguistic, that may contribute to an individual's fluency breakdown, while allowing for the impact of other variables on fluency. Smith and Kelly's model, for example, conceptualizes stuttering as a nonlinear phenomenon, only one component of which is linguistic factors. That is, in attempting to analyze stuttering behavior, it is necessary to acknowledge the multiple "levels" (e.g., linguistic, perceptual, acoustic, sociocultural, etc.) that contribute to it.

In addition to theoretical motivation for the study, the need to more closely examine lexical features of language produced by CWS would appear to be strongly motivated by a growing body of empirical evidence of differences in lexical abilities between CWS and peers. Indeed, Anderson and Conture (2000) noted that the gap between vocabulary knowledge and overall language performance was significantly wider in CWS than the gap observed with normally fluent peers. The primary method used to explore vocabulary use in children's spontaneous language samples has been the calculation of lexical diversity. In the next section, we address some of the issues that have surrounded the use of measures of lexical diversity and their computations.

1.1. Lexical diversity in spontaneous language

Lexical diversity in the spontaneous language of CWS has been a focus of a number of recent investigations (e.g., Ratner & Silverman, 1998; Watkins & Yairi, 1997; Watkins et al., 1999). Historically, the formula used to calculate lexical diversity in child and adult language corpora is the *TTR*, computed by taking the number of unique word roots and dividing by the number of total words in a sample (Miller, 1981). A low *TTR* describes little lexical diversity, while a higher value denotes a language sample composed of a high number of different word roots. A *TTR* of 0.01, for example, could characterize a child repeating the same word 100 times. In contrast, a *TTR* of 1.00 would describe the unnatural situation in which all 100 words of the sample were different. A language sample of syntactically well-formed utterances, therefore, obtains a *TTR* between these two extremes, with a range of content words and the necessary repetition of a small group of function words. A popular alternative measure (Templin, 1957) is *NDW*, which sums the number of unique word roots in a sample of fixed size. The use of *TTR* and *NDW* to gauge diversity of spoken vocabulary in the speech of children who stutter has produced equivocal findings when contrasted with studies that have used alternative measures (e.g., *LR*, Beals & Tabors, 1995) with child populations. For example, Ratner and Silverman (2000) found that the *TTR*s of 15 CWS were not significantly different from those of comparison fluent peers. Conversely, Watkins and Yairi (1997) have used *NDW* to distinguish the expressive language samples of persistent and recovered CWS and their normally fluent peers. Results suggested a slightly elevated *NDW*, relative to normative data, for children who persisted in stuttering for 36 months or more. Similarly, Bonelli et al. (2000), in an analysis of Australian CWS' conversational speech, found pre-treatment values for *NDW* that were well in excess of published norms for the children's chronological ages.

In Ratner and Silverman (2000), measures of *LR* (Beals & Tabors, 1995) were used to augment typical diversity measures calculated over the total sample of words used by the child. In *LR* analyses, the most common words to which children are exposed in the course of early language learning are filtered from the output to yield a proportion of words considered rare or relatively sophisticated in children's conversation.

Either population variation or the measurement metric used in assessing lexical diversity might produce findings that conflict across studies. In particular, there has been recent concern over the validity of various measures of lexical diversity as indicators of the richness of children's lexical output (e.g., Richards & Malvern, 1997; Stokes & Fletcher, 2000; Watkins, Kelly, Harbers, & Hollis, 1995).

1.2. Limitations of TTR and NDW

As noted earlier, *TTR* is the ratio of the number of different words in a sample to the total number of words in the sample. Unfortunately, *TTR* has been shown in a number of analyses to be extremely sensitive to differences in sample sizes; larger samples tend to yield lower *TTR* values (e.g., Hess, Haug, & Landry, 1989; Hess, Sefton, & Landry, 1986; Richards, 1987). Thus, in naturalistic language sampling, variety in the sizes of samples obtained may contaminate derived scores for experimental groups. Additionally, the validity of *TTR* as an indicator of language proficiency or complexity has been challenged: it sometimes fails to discriminate between groups of children who overtly differ in age or diagnostic category (e.g., typically developing versus language impaired, Klee, 1992; Watkins et al., 1995). In the latter cases, use of *NDW*, which is the simple count of different word roots in a fixed size sample, has been shown to be more accurate in discriminating among groups of children with language impairments. However, as Malvern and Richards (1997) note, *NDW* is mathematically confounded with MLU when samples are controlled for length by the number of utterances rather than the number of words (e.g., when 50-*utterance* samples are used, rather than 200-*word* samples). Of course when *NDW* and *TTR* are obtained on samples of the same numbers of words, the resulting values are proportionate to each other; *TTR*s of samples of 200 words each, for example, are simply the *NDW* divided by 200.

It seems that a primary concern related to the use of *TTR* and *NDW* is the issue that samples must be truncated to a common number of words or utterances in order to be compared. It is essential in using these measures to discard potentially informative language sample data to obtain length equivalence across samples. Aside from the fact that we would want to use whole samples to gain as full a picture as possible of a child's language performance, a second issue is the arbitrary nature of selecting the subset of utterances to be analyzed. Investigators have differed in the method used to compute the *NDW* in children's language samples. Some investigators have computed *NDW* based on a standard time period (e.g., Dollaghan et al., 1999; Gavin & Giles, 1996; Robertson & Ellis Weismer, 1999). Others (e.g., Goffman & Leonard, 2000; Klee, 1992; Ratner & Silverman, 2000; Stokes & Fletcher, 2000; Watkins et al., 1995) selected a first or middle subset of utterances or words (e.g., first 100 utterances, middle 100 utterances, etc.) from each sample. Selecting a middle subset, as Templin (1957) did, seems like a reasonable procedure, because selecting from the middle allows the child a "warm-up" period, after which the language obtained from the child presumably would be more representative of everyday language performance. Selecting a particular portion of utterances or words is somewhat problematic, however, given that language samples are generally obtained with a standard set of toys or conversational topics. If only a portion of each sample is analyzed, the children will likely vary in the content of their utterances at that point in the sample. And, given that some play items and topics are simply more well suited for language elicitation than others, subsets of utterances within the sample might vary in the extent to which they adequately represent a child's language abilities. A measure that does not require the limiting of language sample data would, therefore, be more desirable and seemingly more valid for the measurement of children's lexical abilities in conversation.

1.3. Vocd: an alternative

Malvern and Richards (1997) have offered a solution to the problem of using *TTR* for language samples of varying sizes. *Vocd* is a mathematical algorithm applied to *TTR*. As McKee, Malvern, and Richards (2000, p. 323) note:

> The new measure is calculated by, first, randomly sampling words from the transcript to produce a curve of the *TTR* against Tokens for the empirical data. Then the software finds the best fit between this empirical curve and theoretical curves calculated from the model by adjusting the value of a parameter. The parameter, *D*, is shown to be a valid and reliable measure of vocabulary diversity without the problems of sample size found with previous methods.

Vocd allows the input of samples of any size (greater than 50 words), reporting *D* as the measure of lexical diversity. *Vocd* is a relatively new component of the language analysis program *CLAN* (MacWhinney, 2000). The program and the *CLAN* manual can be accessed and downloaded online (http://childes.psy.cmu.edu/) at no charge. The *CLAN* manual describes the *vocd* program and its computational attributes. In order to use *CLAN*'s language sample analysis programs, language samples must first be coded in *CHAT* format, the manual for which is also provided on the *CHILDES* website. To date, the value of using *vocd* to characterize lexical diversity in the speech of young children has been argued primarily from a computational standpoint, rather than from use of the algorithm to describe expressive language profiles of normally developing and communicatively impaired children.

Also conspicuously lacking in most studies of expressive lexical diversity in children's speech are accompanying measures that independently assess the depth and extent of children's vocabulary knowledge. For example, we might expect that children who demonstrate a more restricted array of distinct expressive vocabulary types (with samples obtaining lower *TTR* or *NDW* values) might score more poorly than counterparts on standardized tests of vocabulary comprehension or production. Ratner and Silverman's (2000) study of children's expressive language samples included extensive testing on a number of standardized language tests. Such data permit the evaluation of a number of questions about both the language skills of CWS as well as the relative strengths of various computational algorithms in appraising lexical diversity in children's spoken language samples.

The present study was designed to evaluate the use of *vocd* as a measure of lexical diversity in conversation. Ratner and Silverman (2000) had reported that the CWS, no more than 4 months post-onset, performed significantly differently than their normally fluent peers on a measure of one-word expressive vocabulary, the *Expressive One-Word Picture Vocabulary Test — Revised* (Gardner, 1990). We might speculate that the observed performance difference on a standardized measure of vocabulary might also be evident in the lexical diversity of the CWS' spontaneous language. The results of our previous study, in which *TTR* was used as a measure of lexical diversity, suggested that the groups did not differ on this measure, however. Given the recent concerns, discussed above, about the use of *TTR* as a measure of lexical diversity, the present investigation, involving the same language sample data as reported in Ratner and Silverman (2000), was undertaken to evaluate *vocd* as an alternative, potentially more valid measure. Specifically, the following research questions were posed:

1. To aid in the interpretation of differences in the findings of past research on CWS, what is the relationship between *D* and *TTR*,
   - when whole samples uncontrolled for length are used for both analyses?
   - when a truncated, standard-size sample (50 utterances) is used for *TTR*?
2. As a measure of the concurrent validity of *vocd*, what is the strength of the relationships between *D* and two standardized measures of receptive and expressive one-word vocabulary?
3. How do young children who stutter compare to peers in their lexical diversity, as measured by *D*, as opposed to *TTR*?
4. Given that *vocd* has been advanced as a measure of lexical diversity that is not impacted by differences in sample size, how do whole language samples compare to split-half samples (every other utterance in each sample) in terms of the *D* values obtained?

2. Method

2.1. Participants

Participants in this study were 15 children who stuttered (mean age: 35 months; range: 27–47 months) and 15 normally fluent peers (mean age: 35.67 months; range: 27–47 months). These children had been recruited for a larger study (Miles & Ratner, 2001; Ratner & Silverman, 2000) that examined a number of child and parental behaviors close to stuttering onset. Each of the children in the experimental group was within 4 months (mean = 2.53 months) post-onset of stuttering. The children had an average stuttering frequency of 9.5% (S.D. = 7.2, range = 2.0–25.5%) stuttered words, a category that included sound, syllable, and monosyllabic whole-word repetition, blocks, prolongations, and broken words. The percent stuttered words was calculated from the same language samples as were used for language analyses. None of the children had been seen previously for any speech/language assessment. The CWS and normally fluent peers were pair-matched by gender, age, and maternal level of education. Twelve of the pairs were male; three were female. Each matched cohort of children included two African American children, one child of mixed racial ancestry, and two children from single-parent (mother-only) households. One CWS came from a household in which the mother currently stuttered; no other children had parents who stuttered, although some of the CWS had other relatives who stuttered. None of the children had suspected delays in speech or language. A parent questionnaire completed by all parents revealed no concerns on the part of parents related to their children's hearing. All children and parents spoke a standard dialect of American English and came from middle- to upper-middle-class families (mean level of maternal education: 16 years).

2.2. Language assessment measures

All children participated in an extensive evaluation that included a number of standardized language measures and the collection of a spontaneous language sample. The test battery included the *Peabody Picture Vocabulary Test — Revised* (*PPVT-R*) (Dunn & Dunn, 1981), the *Expressive One-Word Picture Vocabulary Test — Revised* (*EOWPVT-R*) (Gardner, 1990), and two subtests of the *Clinical Evaluation of Language Fundamentals — Preschool* (Wiig, Secord, & Semel, 1992): "Linguistic Concepts" and "Word Structure." Details of standardized test performance other than the two measures (*PPVT-R* and *EOWPVT-R*) related to the present study are not provided herein but can be found in Ratner and Silverman (2000). Spontaneous language samples were elicited while parents engaged in play with a standard set of toys (e.g., building blocks, doctor's kit, play food). Samples ranged in length from 92 to 2329 words (average = 426.9 words; S.D. = 389.2 words). MLU ranged from 2.27 to 4.92 morphemes (average = 3.18 morphemes; S.D. = 0.71 morphemes) for the CWS and ranged from 2.58 to 6.12 morphemes (average = 3.83 morphemes; S.D. = 1.04 morphemes) for their peers. Of interest, only one of the CWS had an MLU of more than 1 S.D. below the mean for children his age (Retherford, 2000). This participant was age 3;11 and had a MLU of 2.48 morphemes. All other participants demonstrated age-appropriate MLU scores. Language samples were entered and coded according to conventions of *CHAT* (MacWhinney, 2000), and both *TTR* and *vocd* language analyses were obtained using *CLAN* (MacWhinney, 2000).

Research questions were addressed using Independent Sample *t*-tests (Question 3), Paired Sample *t*-tests (Question 4) and Pearson's Product Moment Correlations (Questions 1 and 2). Cohen's *d* effect sizes were computed for significant *t*-tests, and values were interpreted as small, moderate, or large, according to Cohen's (1988) guidelines.

3. Results

3.1. Standardized test results

As reported in Ratner and Silverman (2000), both groups performed similarly on the receptive vocabulary measure (*PPVT-R*). Mean standard scores for the two groups were 105.6 (range: 78–121) for the CWS and 110.7 (range: 88–125) for the CWDNS ($t_{(28)} = 1.29$, $P = 0.21$). However, on the *EOWPVT-R* the CWS performed significantly more poorly, as a group, than fluent peers; mean standard scores were 109.4 (range: 89–143) and 120.7 (range: 100–145), respectively ($t_{(28)} = 2.05$, $P = 0.05$; $d = 0.75$, large effect).

3.2. Lexical diversity analyses

Question 1 above asked whether *TTR* and *D* maintained a stable relationship to each other across sample sizes of varying lengths. Our findings suggested that they do not. Figs. 1 and 2 illustrate the relationship between the two measures for full language samples, uncontrolled for length, and for truncated samples of

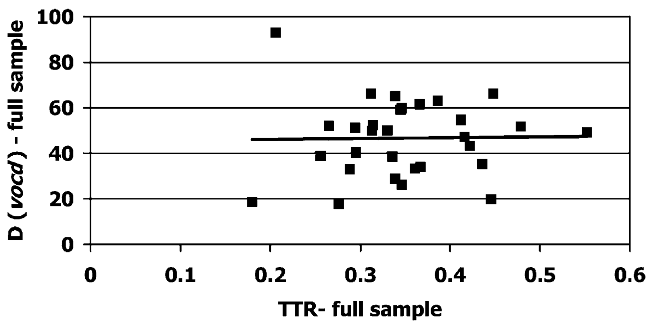

Fig. 1. Relationship between *TTR* values and *vocd* values when whole samples are input ($n = 30$).

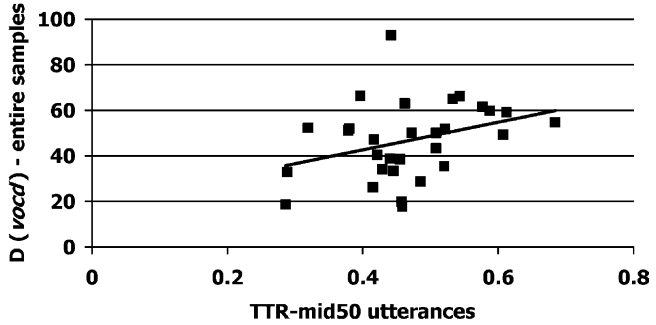

Fig. 2. Relationship between *TTR* values and *vocd* values when the middle 50 utterances are input for *TTR* and the whole sample is input for *vocd* ($n = 30$).

the middle 50 utterances. The correlation between *D* and *TTR* was extremely low ($r = 0.02$, $P = 0.92$) when calculated across full samples but rose substantially, although the relationship was not significant ($r = 0.34$, $P = 0.07$), when a standard 50-utterance corpus was used for *TTR* calculation.

Because *TTR* is known to be sensitive to varying sample sizes, the nonsignificant relationship between *D* and *TTR* when whole language samples are used might be expected. When *TTR* was computed on only the middle 50 utterances as it was intended (see Miller, 1981; Retherford, 2000), and *D* was computed on the entire samples as intended, the relationship between the two measures was not significant, a finding that is somewhat surprising, given that both measures are intended to measure the same construct and are based on the same language sample data.

Our second question asked whether either of the spontaneous vocabulary measures, *TTR* or *vocd*, was related to the children's performance on either receptive or expressive standardized vocabulary tests. Results of this correlational analysis are illustrated in Figs. 3 and 4. Though the relationship between the expressive measure, *EOWPVT-R*, and *D* using whole samples was positive and significant ($r = 0.48$, $P = 0.01$), the relationship between receptive vocabulary and *D* was not significant ($r = 0.33$, $P = 0.08$). However, neither the receptive ($r = -0.31$, $P = 0.09$) nor the expressive ($r = 0.10$, $P = 0.61$) vocabulary score was significantly related to *TTR* using truncated samples. As measures of expressive lexical skill, the most highly correlated scores were of the *EOWPVT-R* and *D*.

Our third question was whether CWS and fluent peers differ in their patterns of lexical diversity, as measured by either *vocd* or *TTR*. Results of this analysis are illustrated in Figs. 5 and 6. Whereas the *vocd* analysis indicated that the lexical diversity of the children who stuttered was significantly less than that of fluent peers ($t_{(28)} = 2.70$, $P = 0.01$, $d = 0.98$, large effect), *TTR* did not differ significantly between groups ($t_{(28)} = 0.41$, $P = 0.69$).

Finally, we were interested in examining the assertion that *D* is in fact relatively stable with variation in sample sizes. In order to assess potential length effects, *vocd* analysis was performed on split halves of each sample, such that for each sample,

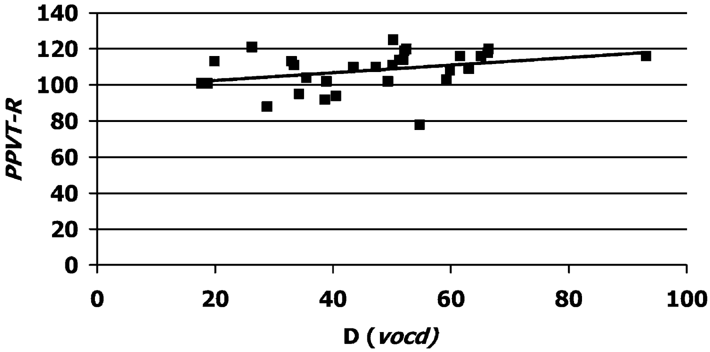

Fig. 3. Relationship between performance on a receptive vocabulary measure (*PPVT-R*) and *vocd* ($n = 30$).

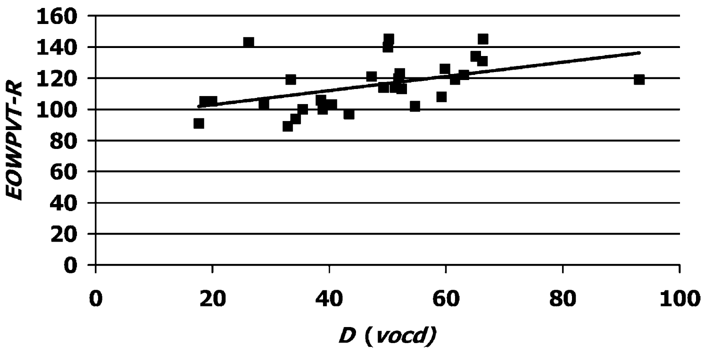

Fig. 4. Relationship between performance on an expressive vocabulary measure (*EOWPVT-R*) and *vocd* ($n = 30$).

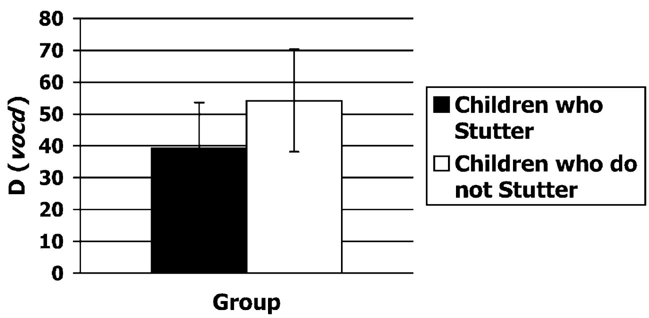

Fig. 5. Comparison of conversational vocabulary as measured by *vocd* (using entire samples) among CWS vs. normally fluent peers ($n = 15$). Standard deviations denoted with error bars.

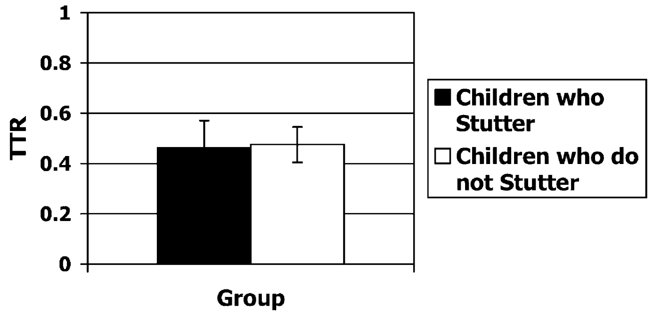

Fig. 6. Comparison of conversational vocabulary as measured by *TTR* (using middle 50 utterances) among CWS vs. normally fluent peers ($n = 15$). Standard deviations denoted with error bars.

every other utterance was omitted from the analysis. Half-sample *D* values (even utterances and odd utterances) were then compared to whole sample *D* values, using two paired-samples *t*-tests. This procedure is identical to that reported by McKee et al. (2000). Data from one CWS were not included in this analysis because his sample size, once split in half, did not meet *vocd*'s sample length requirement of 50 words or greater. Results revealed no significant difference between either the even utterances versus all utterances ($t_{(28)} = 0.29$, $P = 0.77$) or the odd utterances versus all utterances ($t_{(28)} = 1.14$, $P = 0.26$). These results are consistent with those obtained by McKee et al. (2000).

4. Discussion

There is some evidence in the present sample of CWS, evaluated four months or less following onset of their stuttering, of lexical differences from their normally fluent peers in measures of *LR* (see Ratner & Silverman, 2000) and in standardized scores of expressive vocabulary. Additional differences, while not observed through *TTR* measurements (Ratner & Silverman, 2000), were evident in the *vocd* analysis, which has been hypothesized to be a more robust measure of lexical diversity in spoken language. None of the participants in the present study could be described, based on overall language test performance, as having a language disorder; however, it may be the case that the differences in performance observed on measures of *vocd*, the *EOWPVT-R*, and *LR* together point to more subtle differences in language production.

These results are not consistent with those found by Watkins and Yairi (1997), that CWS demonstrate a slightly more diverse lexical repertoire in conversation; potential reasons for the divergent findings may lie either in the heterogeneity of the CWS population or the measures employed to appraise lexical ability. Because the children in the current study showed evidence of diminished lexical performance on three concurrent and converging measures of vocabulary use, we feel somewhat confident in characterizing the current sample of CWS as less proficient in the lexical domain than their fluent peers. Whether or not such proficiency differences can be generalized to the population of children who stutter as a whole is unclear and will require further investigation with larger groups of CWS.

As Anderson and Conture (2000) noted, the accumulating literature suggesting some subtle level of involvement of the lexicon in childhood stuttering, when paired with a small literature suggesting subtle lexical deficits in adults, makes the continued study of lexical functioning in CWS a promising venture. Yaruss and Conture (1996) noted ways in which inefficient lexical and/or phonological access patterns could conceivably produce the behavioral symptoms of stuttering. Future work might target speed and accuracy of lexical access, in addition to other analyses of spoken language style. The power of such lexical analyses to shed light on the underlying nature of stuttering would be heightened by continued emphasis on studying children very near the onset of stuttering symptoms. If lexical patterns observed at stuttering onset can be firmly related to the presenting symptoms of children who stutter, work to strengthen lexical access skills in young children could become an important component of early intervention. Such emphasis would also be consistent with data suggesting that language proficiency is a strong predictor of recovery from early stuttering (Yairi, Ambrose, Paden, & Throneburg, 1996).

The significant relationship between the *D* and a standardized expressive vocabulary measure provides some evidence of concurrent validity for *D* as a measurement of lexical diversity. The fact that the *PPVT-R* was not found to be significantly related to *D* calculations is not particularly surprising, given that conversational vocabulary and one-word vocabulary comprehension represent different aspects of vocabulary knowledge. The relationship between *TTR* and the *EOWPVT-R* was much lower than that for *D* and the *EOWPVT-R*.

Few studies in the literature have tested the concurrent validity of *TTR* by comparing scores with other measures of vocabulary performance. Similar to the present study, Hess, Richie, and Landry (1984), in their study of 6–8-year-old children, found that *TTR* was not significantly related to *PPVT-R* scores. However, in contrast to the present findings, Salzmann (1988) found that among a group of kindergarten children, there was a moderate correlation between *TTR* and the *EOWPVT-R*. One possible reason for this difference in results across studies could be that Salzmann elected to control sample length by the number of words, whereas in the present study, sample length was controlled by the number of utterances. Of interest, however, Hess and colleagues controlled sample length by the number of words in samples and yet did not find the predicted relationship between *TTR* and a standard vocabulary measure. In sum, the findings of these studies in relation to the present study provide conflicting evidence as to the extent to which *TTR* can be validated by standardized one-word vocabulary measures. However, findings of the present study with respect to the relationship between *vocd* and a standardized expressive vocabulary measure, although preliminary, provide some external validation of *vocd*.

Analyses of the effect of sample length, performed by comparing half samples to whole samples, appear to suggest that *D* is not sample size-sensitive. These findings are consistent with those of McKee et al. (2000; cf. Owen & Leonard, 2001). As such, *vocd* would seem to be a reasonable solution to the use of *TTR* and *NDW*, both of which can be sensitive to sample size variation. Moreover, given that *D*, at least in the present study, demonstrated adequate concurrent validity with a standardized measure of expressive vocabulary, it would seem to be a promising measure, both clinically and empirically, for estimating children's conversational vocabulary.

In summary, our current analyses provide one explanation for nonuniform findings of relatively weaker linguistic performance in CWS (e.g., see Nippold, 1990, for a discussion), while suggesting a remedy to the typical research practice of controlling sample sizes. Such a practice is less desirable because language sample data have to be discarded in the process, and because it can fail to take into account situational variability within and across language sample analyses. Future work to link behavioral observations of subtly depressed lexical performance in CWS with analyses of the locus of fluency breakdown in their speech, standardized test performance, and more sensitive measures of lexical access should enable us to further refine the potential contribution of lexical processes to the underlying source or proximal causes of fluency breakdown in early stuttering.

Acknowledgments

We wish to thank Brian Richards for helpful feedback on an earlier version of this manuscript.

References

Anderson, J., & Conture, E. (2000). Language abilities of children who stutter: A preliminary study. *Journal of Fluency Disorders*, *25*, 283–304.

Beals, D., & Tabors, P. (1995). Arboretum, bureaucratic, and carbohydrates: Preschoolers' exposure to rare vocabulary at home. *First Language*, *15*, 57–76.

Bonelli, P., Dixon, M., Ratner, N., & Onslow, M. (2000). Pre- and post-treatment characteristics of adult-child interactions of Lidcombe stuttering program participants. *Clinical Linguistics and Phonetics*, *14*, 427–446.

Byrd, K., & Cooper, E. (1989). Expressive and receptive language skills in stuttering children. *Journal of Fluency Disorders*, *14*, 121–126.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, *42*, 1432–1443.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test — Revised*. Circle Pines, MN: American Guidance Service.

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. F. (1990). *Expressive One-Word Picture Vocabulary Test — Revised*. Novato, CA: Academic Therapy Publications.

Gavin, W. J., & Giles, L. (1996). Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech and Hearing Research, 39,* 1258–1262.

Goffman, L., & Leonard, J. (2000). Growth of language skills in preschool children with specific language impairment: Implications for assessment and intervention. *American Journal of Speech-Language Pathology, 9,* 151–161.

Gordon, P., & Luper, H. (1989). Speech disfluencies in nonstutterers: Syntactic complexity and production task effects. *Journal of Fluency Disorders, 14,* 429–445.

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.

Hess, C. W., Richie, K. P., & Landry, R. G. (1984). Type-token ratio and vocabulary performance. *Psychological Reports, 55,* 51–57.

Hess, C. W., Sefton, K. M., & Landry, R. G. (1986). Sample size and Type-Token Ratios for oral language of preschool children. *Journal of Speech and Hearing Research, 29,* 129–134.

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders, 12*(2), 28–41.

Logan, K., & Conture, E. (1995). Length, grammatical complexity, and rate differences in stuttered and fluent conversational utterances of children who stutter. *Journal of Fluency Disorders, 20,* 35–61.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Erlbaum.

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.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing, 15*(3), 323–338.

McLaughlin, S. F., & Cullinan, W. L. (1989). Disfluencies, utterance length, and linguistic complexity in nonstuttering children. *Journal of Fluency Disorders, 14,* 17–36.

Miles, S., & Ratner, N. (2001). Parental language input to children at stuttering onset. *Journal of Speech, Language, and Hearing Research, 44,* 1116–1130.

Miller, J. (1981). *Assessing language production in children*. Baltimore: University Park Press.

Murray, H., & Reed, C. (1977). Language abilities of preschool stuttering children. *Journal of Fluency Disorders, 2,* 171–176.

Neilson, M., & Neilson, P. (1987). Speech motor control and stuttering: A computational model of adaptive sensory-motor processing. *Speech Communication, 6,* 325–333.

Nippold, M. (1990). Concomitant speech and language disorders in stuttering children: A critique of the literature. *Journal of Speech and Hearing Disorders, 55,* 51–60.

Owen, A. J., & Leonard, L. B. (2001). *Lexical diversity in the speech of normally developing children and children with specific language impairment*. Paper presented at the Symposium on Research in Child Language Disorders, Madison, WI.

Pearl, S. Z., & Bernthal, J. E. (1980). The effect of grammatical complexity upon dysfluency behavior of nonstuttering preschool children. *Journal of Fluency Disorders, 5,* 55–68.

Ratner, N. (1997). Stuttering: A psycholinguistic perspective. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2nd ed., pp. 99–127). Needham, MA: Allyn & Bacon.

Ratner, N., & Sih, 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.

Ratner, N., & Silverman, S. (1998). *Lexical diversity of children's speech at stuttering onset*. Poster presented to the American Speech-Language Hearing Association, San Antonio, TX.

Ratner, N., & Silverman, S. (2000). Parental perceptions of children's communicative development at stuttering onset. *Journal of Speech, Language, and Hearing Research, 43,* 1252–1263.

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.

Richards, B. J., & Malvern, D. D. (1997). *Quantifying lexical diversity in the study of language development*. Reading: New Bulmershe Papers, University of Reading.

Robertson, S. B., & Ellis Weismer, S. (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech, Language, and Hearing Research, 42*, 1234–1248.

Ryan, B. (1992). Articulation, language, rate and fluency characteristics of stuttering and nonstuttering preschool children. *Journal of Speech and Hearing Research, 35*, 333–342.

Ryan, B. (2000). A longitudinal study of articulation, language, rate, and fluency of 22 preschool children who stutter. *Journal of Fluency Disorders, 26*(2), 107–127.

Salzmann, M. (1988). *An investigation of the Type-Token Ratio and the Expressive One-Word Vocabulary Test*. Unpublished MA thesis, Ball State University, Muncie, IN.

Silverman, S., & Ratner, N. (1997). Syntactic complexity, fluency and accuracy of sentence imitation in adolescents. *Journal of Speech, Language, and Hearing Research, 40*, 95–106.

Smith, A., & Kelly, E. (1997). Stuttering: A dynamic, multifactoral model. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2nd ed., pp. 204–217). Needham, MA: Allyn & Bacon.

Starkweather, C. W. (1987). *Fluency and stuttering*. Englewood Cliffs, NJ: Prentice Hall.

Stokes, S. F., & Fletcher, P. (2000). Lexical diversity and productivity in Cantonese-speaking children with specific language impairment. *International Journal of Language and Communication Disorders, 35*, 527–541.

Templin, M. (1957). *Certain language skills in children*. Minneapolis: University of Minnesota Press.

Wall, M. J., & Myers, F. L. (1995). Theories of stuttering and therapeutic implications. *Clinical management of childhood stuttering* (2nd ed.). Austin, TX: Pro-Ed.

Watkins, R., Kelly, D., Harbers, H., & 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.

Weiss, A. L., & Zebrowski, P. M. (1992). Disfluencies in the conversations of young children who stutter: Some answers about questions. *Journal of Speech and Hearing Research, 35*, 1230–1238.

Westby, C. (1979). Language performance of stuttering and nonstuttering children. *Journal of Communication Disorders, 12*, 133–145.

Wiig, E., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals — Preschool*. San Antonio, TX: Psychological Corporation.

Yairi, E., Ambrose, N., Paden, E., & Throneburg, R. (1996). Predictive factors of persistence and recovery: Pathways of childhood stuttering. *Journal of Communication Disorders, 29*, 51–77.

Yaruss, J. S., & Conture, E. G. (1996). Stuttering and phonological disorders in children: Examination of the Covert Repair Hypothesis. *Journal of Speech and Hearing Research, 39*, 349–364.

CONTINUING EDUCATION

Measuring lexical diversity in children who stutter: application of *vocd*

QUESTIONS

1. The purpose of the present study was to:
   a. evaluate the use of *vocd* as a measure of lexical diversity for use with children who stutter

   b. compare younger and older children's conversational vocabulary abilities using *vocd*
   c. use *vocd* to diagnose language impairment in young children who stutter
   d. evaluate the use of *vocd* for the diagnosis of stuttering in young children
   e. develop a new standardized measure of vocabulary knowledge for use with young children
2. The conversational vocabulary abilities of children who stutter have been assessed using all of the following measures, except:
   a. Type-Token Ratio
   b. Number of Different Words
   c. Articulatory speaking rate
   d. Lexical Rarity measures
   e. None of the above; they are all measures of conversational vocabulary abilities
3. What is *vocd*?
   a. A standardized measure of receptive vocabulary
   b. A standardized measure of word retrieval abilities
   c. A language analysis procedure for measuring lexical diversity in language samples of varying lengths
   d. A program of *CHILDES* that computes lexical diversity while taking syntactic complexity into account
   e. Both c and d
4. All of the children selected for the experimental group of the present study:
   a. were previously diagnosed as having a language impairment
   b. had begun stuttering no more than 4 months before participating in the study
   c. had recovered from stuttering no more than 4 months before participating in the study
   d. had a family history of stuttering
   e. were at least 5 years old
5. Findings of the present study were that:
   a. *D* values were significantly related to performance on the *EOWPVT*
   b. *D* values were significantly related to *TTR* values
   c. CWS and fluent peers differed in their lexical diversity as measured by *D*
   d. CWS and fluent peers differed in their lexical diversity as measured by *TTR*
   e. Both a and c