Syracuse University

SURFACE

Communication Sciences and Disorders

College of Arts and Sciences

2014

Speech disfluencies of preschool-age children who do and do not stutter

Victoria Tumanova
Syracuse University, vtumanov@syr.edu

Edward G. Conture
Vanderbilt University

Warren Lambert
Vanderbilt University

Tedra A. Walden
Vanderbilt University

Follow this and additional works at: https://surface.syr.edu/csd

Part of the Communication Sciences and Disorders Commons

Recommended Citation

Speech disfluencies of preschool-age children who do and do not stutter

Victoria Tumanova[a,*], **Edward G. Conture**[b,1], **E. Warren Lambert**[c,2], and **Tedra A. Walden**[d,3]

Edward G. Conture: edward.g.conture@vanderbilt.edu; E. Warren Lambert: warren.lambert@vanderbilt.edu; Tedra A. Walden: tedra.walden@vanderbilt.edu

[a]Department of Communication Sciences and Disorders, Syracuse University, 621 Skytop Road, Syracuse, NY 13244, United States

[b]Department of Hearing and Speech Sciences, Vanderbilt University, 1215 21st Avenue South, Nashville, TN 37232, United States

[c]Department of Special Education, Vanderbilt University, 203 One Magnolia Circle, Nashville, TN 37232, United States

[d]Department of Psychology & Human Development, Vanderbilt University, 230 Appleton Place, Nashville, TN 37232, United States

Abstract

Purpose—The goals of the present study were to investigate whether (1) the speech disfluencies of preschool-age children are normally distributed; (2) preschool-age children who do (CWS) and do not stutter (CWNS) differ in terms of non-stuttered disfluencies; (3) age, gender, and speech-language ability affect the number and type of disfluencies children produce; and (4) parents' expressed concern that their child stutters is associated with examiners' judgments of stuttered disfluency.

Method—Four hundred and seventy two children participated, of which 228 were CWS (56 girls), and 244 CWNS (119 girls). Participants provided conversational speech samples that were analyzed for frequency of occurrence of (a) stuttered disfluencies, (b) non-stuttered disfluencies, and (c) total disfluencies.

Results—Results indicated that the underlying distributions of preschool-age children's stuttered and non-stuttered disfluency counts followed a negative binomial distribution (i.e., were not normal), with more children "piling up" at the low end [none or few disfluencies] and fewer children scoring in the upper [more severe stuttering] end of the distribution. Findings also indicated that non-stuttered disfluencies significantly predicted CWS/CWNS talker group classification, information that may be helpful to augment, but not supplant, talker group classification criteria based on stuttered disfluencies. Moreover, expressed parental concern about stuttering was strongly associated with frequency of stuttered disfluencies.

*Corresponding author. Tel.: +1 315 443 9640; fax: +1 315 443 4413. victoria.tumanova@gmail.com (V. Tumanova).
[1]Tel.: +1 615 322 8780; fax: +1 615 936 6914.
[2]Tel.: +1 615 322 7411.
[3]Tel.: +1 615 322 8346.

Conclusion—Findings suggest that the entirety of preschool-age CWS' speech disfluencies – non-stuttered as well as stuttered – differs from that of their CWNS peers and that because these disfluencies are not normally distributed statistical analyses assuming normality of distribution are not the most appropriate means to assess these differences. In addition, certain "third-order" variables (e.g., gender) appear to impact frequency of children's disfluencies and expressed parental concerns about stuttering are meaningfully related to examiners' judgments of stuttered disfluencies.

Learning outcomes—The reader will learn about differences in speech disfluencies of preschool-age children who do and do not stutter. The reader will learn whether age, gender and speech-language ability affect the number and type of disfluencies children produce. The reader will learn whether parental concern about stuttering is associated with examiners' judgments of stuttering.

Keywords

1. Introduction

During the late 1950s, Johnson et al. (1959) provided normative data regarding the speech disfluencies of children who do and do not stutter. These researchers obtained their data from assessments of audio recordings of children's speech disfluencies. Since then, several others based on similar recordings of speakers of English (e.g., Ambrose & Yairi, 1999; Pellowski & Conture, 2002; Yaruss, LaSalle, & Conture, 1998) and speakers of other languages (e.g., Boey, Wuyts, Van de Heyning, Bodt, & Heylen, 2007; Carlo & Watson, 2003; Martins & Andrade, 2008; Natke, Sandrieser, Pietrowsky, & Kalveram, 2006), have contributed data to the foundation laid down by Johnson and colleagues in the 1950s. Combined, these empirical investigations, studied 908 children who stutter (CWS) and 258 children who do not stutter (CWNS). Although the nature of the samples differed (e.g., some involved the child talking to an experimenter, others the child talking to a caregiver, and some collected the data within research whereas others within a clinical setting), this accumulated dataset represents one of the largest repositories of information presently available regarding the speech disfluencies of CWS and CWNS.

There are, however, some issues relating to this body of knowledge that bear further consideration. First, there is the issue of how underlying characteristics of stuttered (i.e., sound-syllable and monosyllabic whole-word repetitions and sound prolongations) and non-stuttered (i.e., interjections, phrase repetitions and revisions) disfluencies may impact data analysis. For example, are the distributions of such disfluencies Gaussian or normal? Second, is the issue of whether, in addition to stuttered disfluencies, "non-stuttered," "other" or "normal" disfluencies are salient to our understanding and/or classification of developmental stuttering in preschool-age children. Third, is the issue of misattribution of effect, that is, do third-order variables (e.g., age, gender or speech-language status) confound our understanding of between-group differences in speech disfluency. Fourth, is the issue of whether there is an association between parents/caregivers' expressed reports of concern that

their child is or is suspected to be stuttering and examiners' measurement of the child's instances of stuttered disfluencies? Below, we briefly examine each of these issues.

The first issue, the distribution of speech disfluencies, has received little attention in data analyses, with a few exceptions. For example, Johnson, Darley, and Spriestersbach (1963) noted that the frequency distributions of speech disfluencies "are considerably skewed or "long-tailed in one direction" with "piling up of scores toward the low end of the distribution" (p. 252). Similar descriptions were also reported by Davis (1939) and Jones, Onslow, Packman, and Gebski (2006). Johnson and colleagues further speculated that from such distributions "we may draw the generalization that there are more relatively mild than relatively severe stutterers" (p. 252). Interestingly, however, researchers assessing between-group differences in speech fluency (e.g., Yaruss, LaSalle, et al., 1998; Yaruss, Max, Newman, & Campbell, 1998) have typically employed parametric inferential statistical analyses that assume normality of distribution (e.g., analysis of variance, *t*-tests, etc.). Unfortunately, despite the observations of Johnson and colleagues, as well as Davis and others, there is little empirical evidence in the literature that the underlying distributions of reported speech disfluencies (e.g., stuttered disfluencies, non-stuttered disfluencies and so forth) are normally distributed. If the distributions of (non)stuttered disfluencies assume a non-normal or non-Gaussian form (e.g., strong positive skew), then the use of parametric inferential statistics may be problematic. If the assumption of normality cannot be met, then the assumption of ordinary least squares regression or analysis of variance is violated, possibly leading to the rejection of the null hypothesis when in fact it is true. If such violation is the case, it leads to the suggestion that researchers' consider employing analytical statistical models that better fit the data's actual distribution.

A second question concerns the frequency of stuttered disfluencies and non-stuttered or normal disfluencies exhibited by children who do and do not stutter. Many studies of developmental stuttering, and reasonably so, have classified the two talker groups based on frequency of instances of "stuttering" (e.g., Ambrose & Yairi, 1999; Anderson & Conture, 2001; Logan & LaSalle, 1999; Sawyer & Yairi, 2006; Watkins & Yairi, 1997). It should be noted that that some differences do exist across various studies in the way stuttered disfluencies are described as well as what constitutes a stuttered disfluency (for further review, see Einarsdottir & Ingham, 2005). At present, however, some have classified children as stuttering if the child exhibits 3% or greater stuttered disfluencies in their conversational speech sample (e.g., Conture, 2001; Yairi & Ambrose, 2005). Similarly, Boey et al. (2007), based on a large sample of Dutch-speaking children (*n* = 772), reported that the "3% rule" has high specificity (true negative CWNS classifications) and high sensitivity (true positive CWS classifications). However, to the present writers' knowledge, specificity and sensitivity of the "3% rule" have never been assessed in a large sample of English-speaking children.

Although frequency of stuttered disfluencies is often used to diagnose and classify stuttering in children, there is less certainty regarding the salience of "non-stuttered," "other," or "normal" disfluencies to the diagnosis and/or understanding of developmental stuttering. Some studies have reported that CWS produce significantly more non-stuttered disfluencies than CWNS (Ambrose & Yairi, 1999; Johnson et al., 1959; Yairi & Ambrose, 2005)

whereas others did not find any significant difference (Logan, 2003; Pellowski & Conture, 2002; Yairi & Lewis, 1984). One may ask, therefore, whether non-stuttered speech disfluencies of CWS objectively differentiate the two talker groups. If they do differentiate the two talker groups, it would suggest that the entirety of CWS's speech disfluencies, not just the stuttered aspects, differ from typically developing children, at least in terms of frequency of occurrence.

Certainly, previous empirical findings indicate that CWS produce non-stuttered disfluencies; however, these findings are seldom discussed in detail (cf. Ambrose & Yairi, 1999; Pellowski & Conture, 2002). Some authors have also suggested that frequency of total disfluencies (i.e., stuttered plus non-stuttered) provides a reasonable criterion for talker group classification (Adams, 1977). Although the use of total disfluency as criterion for talker-group classification does bring non-stuttered disfluencies under the tent of decisions involved with talker group (CWS vs. CWNS) classification criteria, this criterion is confounded by its inclusion of stuttered disfluencies, the latter shown to significantly distinguish between children who do and do not stutter (e.g., Boey et al., 2007). Nevertheless, Adams' suggestion highlights the possibility that measures besides instances of stuttered disfluency may have diagnostic salience. This possibility raises the question of whether non-stuttered speech disfluencies may augment clinicians' as well as researchers' attempts to develop a data-based diagnosis of developmental stuttering.

A third issue is the potential misattribution of effect. Specifically, when studying possible differences between CWS and CWNS on a particular variable (e.g., frequency of disfluencies during conversational speech), other possible predictors coexist, for example, age, gender, or expressive language abilities. Researchers have often dealt with this issue by matching the two talker groups (i.e., CWS and. CWNS) for age, gender, speech-language abilities, etc. before assessing between-group differences in speech fluency. However, this matching procedure does not necessarily indicate whether, for example, a variable such as chronological age impacts the actual reported between-group (i.e., CWS vs. CWNS) differences in frequency of speech disfluencies, stuttered or otherwise. One way to address this issue is to statistically model potentially confounding variables as covariates. This model-based approach has an advantage over matching talker groups for possible confounds (e.g., age) because it (a) allows the experimenter to obtain representative samples of both talker groups more closely reflective of the natural variation in these variables and, more importantly, and (b) assess whether such variables (e.g., gender) actually impact reported between-group differences in speech disfluencies.

In the present study, and based on review of empirical studies of speech disfluencies in young children, we selected three variables commonly matched or considered when assessing between-group differences: age, gender, and speech-language abilities. These three variables were covariates in our statistical models/data analyses of preschool-age children's speech disfluencies. Certainly, these are not the only possible covariates, but they are three of the most common variables investigators have reported considering when assessing group differences between preschool-age CWS and CWNS. Immediately below we briefly review the possible association of each of these three variables and childhood stuttering.

Regarding the chronological age of preschool-age CWS, it should be noted that most if not all standardized speech-language tests are age-normed. Further, experience with stuttering (i.e., time since onset) in young children is intimately connected to chronological age (e.g., Pellowski & Conture, 2002), with some tests used to assess childhood stuttering, for example, the KiddyCAT, apparently being sensitive to chronological age (e.g., Clark, Conture, Frankel, & Walden, 2012). Indeed, frequency of different disfluency types may vary with age and differ between young and older children (e.g., Davis, 1939; DeJoy & Gregory, 1985; Yairi & Clifton, 1972). Whether chronological age impacts between-group differences in stuttered and non-stuttered disfluencies remains an open empirical question.

With regard to the gender of preschool-age CWS, there is considerable evidence that the prevalence of stuttering is greater in males than females (e.g., Bloodstein & Bernstein Ratner, 2008), and that males are also more at risk for persistence (Yairi & Ambrose, 1992; Yairi & Ambrose, 2005; Yairi, Ambrose, Paden, & Throneburg, 1996). In view of this gender difference among CWS, it seems important to better understand whether gender impacts between-group differences in stuttered and non-stuttered disfluencies, as well as within-group differences. Based on their findings, Johnson et al. (1959) suggest that gender does not impact these between- and within-group differences, but to the present authors' knowledge this issue has not been empirically replicated, especially with large samples of both preschool-age CWS and their CWNS peers.

It is known that speech and language abilities develop with age and that stuttering for many children begins during the time of rapid language growth between the 2.5 and 5 years of age (e.g., Bloodstein & Bernstein Ratner, 2008). Furthermore, there is some evidence of between group-differences (CWS vs. CWNS) in articulation and/or phonological disorder (e.g., Blood, Ridenour, Qualls, & Hammer, 2003; cf. Clark et al., 2013). Likewise, meta-analytical findings suggested that CWS scored significantly lower than CWNS on norm-referenced measures of overall language as well as receptive and expressive vocabulary (Ntourou, Conture, & Lipsey, 2011). Furthermore, preschool-age CWS' tendency to stutter on utterance-initial words increases with sentence length, a finding suggesting that language/linguistic processes contribute to the frequency/locus of stuttering (Richels, Buhr, Conture, & Ntourou, 2010). Indeed, language growth and complexity appears to be associated with children's speech disfluency (e.g., Yaruss, Newman, & Flora, 1999). It has also been shown that children with histories of specific expressive language impairment are significantly more disfluent than their peers with normal language development (Boscolo, Bernstein Ratner, & Rescorla, 2002; Hall, Yamashita, & Aram, 1993; cf. Nippold, 2012). It is not clear, however, whether speech and language abilities, as well as age, affect frequency and type of disfluencies in typically developing children who do and do not stutter.

A fourth issue is the association parents/caregivers' expressed concern that their child is or is suspected to be stuttering and examiners' judgment of frequency of children's stuttered disfluencies. While both measures involve perceptual judgments and are, to greater or lesser degrees, somewhat subjective, examiner judgment is often considered relatively objective and parental concern, while certainly a reality with regard to childhood stuttering, seemingly less so (although some, for example, Yairi & Ambrose, 2005, have used an ordinal scale for parents to judge their level of concern). However, to the present writers' understanding,

there has been little attempt to empirically assess the nature of the association between parents' expressed concern and examiners' judgment of stuttered disfluencies. Although it might be assumed that some degree of association exists, this remains an open empirical question. Knowing that these two variables – examiners judgment of stuttering and parental expressed concern about their child's stuttering – are associated would seem to lend a degree of demonstrated concurrent validity to both measures, something that appears presently lacking.

Overall, some basic variables associated with early childhood speech disfluencies (i.e., distribution of stuttered and non-stuttered disfluencies), as well as the impact of related variables on such disfluencies have received relatively little empirical attention. As a result, our understanding of salient variables associated with preschool-age children's speech fluency is less than comprehensive. Without such understanding, we would seem to be less than aware of how these variables impact our assessment of children's speech disfluencies and/or the diagnosis of childhood stuttering.

Therefore, the present study addressed these issues by examining childhood disfluencies in a large (n = 472) sample of preschool-age children. These data were collected in a consistent manner using a standard battery of testing procedures and measures commonly used by practicing clinicians as well as researchers who study childhood stuttering.

To assess preschool-age CWS and CWNS's speech disfluencies, four questions were investigated. The first question asked whether the stuttered, non-stuttered and total speech disfluencies of preschool-age children who stutter are normally distributed (a question essential to the selection of analytic statistical models). It was hypothesized that preschool-age children's speech disfluencies are not normally distributed, rather that they are positively skewed with more children "piling up" at the low end of the distribution (none or few disfluencies) and fewer children scoring in the upper (more severe stuttering) end of the distribution.

The second question asked whether preschool-age CWS and CWNS differ in terms of stuttered and non-stuttered disfluencies and whether those two variables possess strong classification capacity, thus being useful for differentiating preschool-age CWS from their CWNS peers. We hypothesized that the entirety of CWS's speech, not merely their stuttered disfluencies, is more disfluent than that of CWNS. Specifically, CWS, when compared to CWNS, were hypothesized to exhibit significantly more stuttered, non-stuttered disfluencies and total disfluencies, and that such speech disfluencies would significantly predict talker group membership.

The third question asked whether age, gender, and speech-language ability impact young children's stuttered, non-stuttered, and total disfluencies. We hypothesized that children who have lower speech and language scores will exhibit more disfluencies than those who have higher speech and language scores.

The fourth question related to the association of parental expressed concern that their child stutters and examiners' judgments of stuttering, a particularly salient issue given the common use of "parental concern" to categorize children as stuttering. We hypothesized that

there would be a strong association between parents' expressed concern that their children stutter or are suspected to be stuttering and the frequency of examiner-judged stuttered disfluencies.

2. Method

2.1. Participants

Participants were 228 monolingual, English speaking children who stutter (56 girls and 172 boys, age: $M = 47.47$ months, $SD = 8.94$, range = 31–71 months), and 244 children who do not stutter (119 girls and 125 boys, age: $M = 50.47$ months, $SD = 9.59$, range = 36–71 months),[4] with time since onset freely varying within our sample of preschool-age CWS.

All participants were part of an ongoing series of empirical studies of linguistic and emotional contributions to developmental stuttering conducted as a part of Vanderbilt University's Developmental Stuttering Project (e.g., Arnold, Conture, Walden, & Key, 2011; Choi, Conture, Walden, Lambert, & Tumanova, 2013; Clark et al., 2012; Johnson, Conture, & Walden, 2012; Ntourou, Conture, & Walden, 2013; Richels et al., 2010; Walden et al., 2012). All were paid volunteers whose parents either learned about the study from an advertisement in a monthly parent magazine circulated throughout Middle Tennessee, an e-mail advertisement sent to Vanderbilt University employees, or were referred to the Vanderbilt Bill Wilkerson Hearing and Speech Center for an evaluation. The study procedures were approved by the Vanderbilt University Institutional Review Board. Informed consent by parents and verbal assent by children were obtained.

The Hollingshead Four-Factor Index of Social Position (Hollingshead, 1975) was used in the present study to provide a descriptive measure of participants' socioeconomic status (SES). This index takes into account both parents' educational levels, occupation, and marital status, based on self report. Computed scores ranged from 8 to 66, with a higher score indicating a higher socioeconomic status.

2.2. Measurement of speech fluency

Measurement of participants' speech fluency was based on a 300-word conversational speech sample, obtained during free play between the child and the examiner, and scores on the Stuttering Severity Instrument-3 (SSI-3; Riley, 1994). Scores on the SSI-3 were based on one continuous 300-word conversational speech sample. All disfluency and word counts were obtained in real-time with the examiner noting the disfluent and fluent words on a disfluency count sheet (Conture, 2001) while playing and conversing with the child.

Present study guidelines for assessing speech disfluencies were such that only one disfluency type (e.g., sound/syllable repetition) could be applied to a single word. If two or more stuttered disfluencies (for examples, see below) occurred on the same word (e.g., disfluency cluster "sound prolongation + sound/syllable repetition"), only one instance of stuttered disfluency, that is, the first disfluency to occur on the word, was documented/

---

[4]Apparent between-group difference in gender as well as other relevant variables (e.g., age) will be accounted for in statistical model presented in Section 3.

counted for. Phrase repetitions or revisions (which are classified in this study as non-stuttered/normal disfluencies; for examples, see below) occur on units larger than single words. Thus, if a stuttered and a non-stuttered disfluency occurred within the same phrase (e.g., a sound prolongation on one word of phrase revision), both were counted (see Yaruss, 1998a,b). All examiner-child interactions were audio-video recorded for several purposes, including inter- and intra-judge measurement reliability, to be described below.

2.3. Classification and inclusion criteria

All participants' speech-language and hearing abilities were assessed using standardized measures. In particular, the "Sounds in Words" subtest of the Goldman–Fristoe Test of Articulation-2 (GFTA-2; Goldman & Fristoe, 2000) assessed children's articulation. Receptive vocabulary was measured using the Peabody Picture Vocabulary Test-Third Edition (PPVT-4; Dunn & Dunn, 2007). Expressive vocabulary was measured using the Expressive Vocabulary Test (EVT-2; Williams, 2007). Receptive and expressive language abilities of the participants were evaluated using the Test of Early Language Development-3 (TELD-3; Hresko, Reid, & Hammill, 1999). In addition, all participants received a bilateral pure tone hearing screening to rule out hearing impairments.

Participants were assigned to the CWS group if they (a) exhibited three or more stuttered disfluencies (i.e., sound/syllable repetitions, sound prolongations, or monosyllabic whole-word repetitions) per 100 words of conversational speech (Conture, 2001; Yaruss, 1998a,b) based on a 300-word speech sample, and (b) scored 11 or greater (i.e., severity of at least "mild") on the SSI-3 (Riley, 1994).[5] Participants were classified as CWNS if they (a) exhibited two or fewer stuttered disfluencies per 100 words of conversational speech based on a 300-word sample, and (b) scored 10 or lower on the SSI-3.

2.4. Procedures

Data collection for all participants consisted of a parent interview, wherein information was obtained regarding the family's SES, history of speech-language and fluency disorders, as well as concerns about children's speech-language abilities (for further details pertaining to this interview process, see Conture, 2001; Richels & Conture, 2010). While one examiner conducted the parent interview, another examiner talked with the child during free play, taking the "on-line" disfluency count, from which measures of speech fluency were obtained. Participants were then given a series of standardized speech and language tests in the following fixed order: GFTA-2, PPVT-4, EVT-2, and TELD-3, a procedure, the authors have found, to maximize the chances that the greatest number of preschool-age children will successfully complete all such testing. Standardized testing was followed by the administration of the KiddyCAT (Clark et al., 2012; Vanryckeghem & Brutten, 2007) and bilateral pure tone hearing screenings. Audiometric equipment was routinely calibrated.

Testing of participants was conducted in a controlled laboratory environment as part of a pre-experimental diagnosis/screening to determine inclusion/exclusion for subsequent

---

[5]Although expressed parental concern about stuttering was not used in present talker-group classification, a hypothesis regarding parental concern and frequency of stuttered disfluencies was tested in this study, with findings presented in the results section.

experimental research (e.g., Arnold et al., 2011; Byrd, Conture, & Ohde, 2007; Johnson et al., 2010; Walden et al., 2012).

2.4.1. Expressed parental concern—As described above all parents who participated in this study ($n = 472$) were asked a series of questions about their child's development, including possible concerns with stuttering. It will be recalled that expressed parental concern about stuttering was *not* used for talker group classification, but only to address hypothesis 4 in order to assess the association between parent concern and examiner judgment of speech disfluency. This parental judgment was obtained during their initial contact, by means of telephone and/or email, with our research team. Their affirmative/ negative response was recorded and confirmed again at the time of testing.

2.5. Description of dependent variables

Dependent measures in this study were as follows: (a) number of stuttered disfluencies (SDs), (b) number of non-stuttered disfluencies (NSDs) and (c) number of total disfluencies or (a) + (b) per 300 words of conversational speech.

2.5.1. Stuttered disfluencies—The following disfluency types were considered to be stuttered: (a) sound-syllable repetition (SSR), ("b-but", "le-le-lemon"); (b) monosyllabic whole-word repetition (WWR) ("I-I-I", "my-my-my"); and (c) audible and inaudible sound prolongations (SP) ("mm-mine", "ssss-some", "pa—per").

2.5.2. Non-stuttered ("normal") disfluencies—The following disfluency types were considered to be non-stuttered (or "normal" disfluencies): (a) phrase repetitions (PR) ("I want – I want a cookie"); (b) revisions (REV) ("He went – They went to school"); and (c) interjections (INT) ("uhm").

2.6. Measurement reliability for identification of disfluencies

To assess *inter-judge* measurement reliability, often called inter-rater reliability, for identification of stuttered and non-stuttered disfluencies as well as fluent words in children's speech, the frequency of both was recalculated for 32 children (i.e., 18 CWS and 14 CWNS). Four examiners independently re-evaluated the speech samples by taking a disfluency count in real time while watching a video recording of the previously conducted speech assessment. The samples for re-evaluation were selected at random from each group of preschool-age participants (CWS and CWNS). Reliability of measurement between the original and recalculated data was assessed by calculating intra-class correlation coefficients (ICC; McGraw & Wong, 1996; Shrout & Fleiss, 1979).

Inter-judge reliability ranged from (a) .95 to .97 ($M = .96$), with average ICC measures of . 989, $p < .001$, for identification of stuttered disfluencies; (b) .82 to .89 ($M = .86$), with average measures of .95, $p < .001$, for identification of non-stuttered disfluencies; and (c) .94 to .97 ($M = .96$), with average measures of .98, $p < .001$, for identification of total disfluencies. The above ICC reliability values exceed the popular criterion of .7 (Yoder & Symons, 2010).

To assess *intra-judge* reliability, each of the four examiners re-evaluated disfluency counts of 11 children (*M* = 6 CWS; *M* = 5 CWNS) they had previously completed. Both the inter-judge and intra-judge reliability disfluency counts were taken in real time while watching the video recording of the child-clinician conversation. The time between the first and the second count was at least 3 months. ICCs ranged from .95 to .99 (*M* = .97) for identification of SD, from .8 to .96 (*M* = .93) for identification of NSD, and from .97 to .98 (*M* = .97) for identification of TD.

2.7. Data analysis

To test for the normality of the distribution of speech disfluencies, the present authors used a Shapiro–Wilk test of normality (Shapiro & Wilk, 1965) and inspected distributions with histograms. A histogram for each dependent variable (i.e., total, stuttered, and non-stuttered disfluencies) was plotted, and descriptive statistics were calculated (mean, standard deviation, variance, skewness and kurtosis).

To assess between-group differences (i.e., CWS vs. CWNS) for frequency of stuttered and non-stuttered disfluencies, a generalized linear regression model (Nelder & Wedderburn, 1972) was estimated. This model was chosen because it allows for analysis of data that do not fit a normal distribution. "Generalized" means that various distributions can be chosen, such as binary, Poisson, or negative binomial if the distribution of a dependent variable is not normal. "Negative binomial" refers to a Poisson regression with overdispersion (e.g., a long right-hand tail) and is often used because many counts of events may be more dispersed than the traditional Poisson (Gardner, Mulvey, & Shaw, 1995). Generalized models are provided in various commonly used software packages (e.g., SPSS, SAS, Stata, R) with a statistical basis for such models given in many sources, such as the Hardin and Hilbe (2003) monograph.

To assess whether participants' age, gender and speech-language abilities influenced the frequency of their speech disfluencies, these categorical or continuous independent variables were entered as covariates in the generalized regression model for each dependent variable. Software employed was SPSS-19 "Generalized Linear Models." The model tested for the main effects of each covariate and also for the interaction of group by gender as described below.

To assess whether parental concern about children's stuttering is associated with examiner's judgment of stuttering we employed a logistic regression analysis.

3. Results

Analyses of descriptive data/group characteristics are reported first, followed by statistical tests of each hypothesis.

3.1. Descriptive analyses of the data: group differences in age, gender and speech-language abilities

Table 1 provides descriptive statistics for each talker group for language variables and age, all of which were normally distributed. Normal distributions are common for standardized

tests with many items. Multivariate ANOVA was performed to assess between-group differences on each variable.

Results indicated that preschool-age CWS, when compared to preschool-age CWNS, show significantly lower scores on PPVT ($F = 11.71$, df = 1, $p = .001$), EVT ($F = 11.96$, df = 1, $p = .001$), and TELD receptive subtest standard scores ($F = 13.47$, df = 1, $p < .001$). There was also a significant difference in age ($F = 12.26$, df = 1, $p = .001$) with CWS being younger than CWNS, and our sample included 1.4 times more male CWS ($n = 172$) than male CWNS ($n = 125$).

These differences between preschool-age CWS and their CWNS peers give these variables potential leverage to influence measures of speech disfluency. As mentioned above, to control for possible effects of those differences on stuttered and non-stuttered disfluencies, each of these possible confounds was entered in the statistical model as a covariate. There were no significant between-group differences on the GFTA, TELD expressive subtest standard scores or SES.

3.2. Hypothesis 1: non-normality of distribution of speech disfluencies

Table 2 provides descriptive statistics (percentiles) for both talker groups for all dependent variables (i.e., stuttered, non-stuttered and total disfluencies). Results of the Shapiro–Wilk test of normality indicated that the distributions for all three variables were non-normally distributed. The statistics for distribution of *stuttered* disfluencies were as follows: $W = .954$, df = 244, $p < .0001$ for CWNS and $W = .861$, df = 228, $p < .0001$ for CWS, with significance of the Shapiro–Wilk's test indicating non-normality of distributions for both talker groups. The statistics for distribution of *non-stuttered* disfluencies were as follows: $W = .914$, df = 244, $p < .0001$ for CWNS and $W = .945$, df = 228, $p < .0001$ for CWS, also non-normal distributions for both talker groups. The statistics for distribution of *total* disfluencies were as follows: $W = .947$, df = 244, $p < .0001$ for CWNS and $W = .897$, df = 228, $p < .0001$ for CWS, again, non-normal distributions for both talker groups. Consistent with these analytical findings, histograms for each of the three dependent variables (Fig. 1(A)–(C)) show that the data were non-normally distributed. The skewed distributions resembled a Poisson distribution but the variance was excessive (larger than the mean). For this reason a negative binomial distribution was used to model the distributions.

In brief, results of both formal and informal assessment of normality supported hypothesis 1, that is, stuttered, non-stuttered and total disfluencies were not normally distributed. Specifically, the distribution for each of the dependent variables was skewed to the right (positively skewed) indicating that the mass of each of the distributions was concentrated in the lower end of the disfluency continuum with more "mild" disfluencies for CWS and greater fluency for CWNS. The descriptive indices of normality are presented in Table 3.

3.3. Hypotheses 2 and 3: between-group differences on speech disfluencies

Since both the second and the third hypotheses were tested in the same statistical model, results of those analyses are reported together. Again, as described above, generalized linear regression analysis – a procedure that can be used for various distributions of dependent

variables – was employed to assess between-group differences (CWS vs. CWNS) in the frequency of stuttered, non-stuttered and total disfluencies during children's conversational speech.

To test the hypothesis of whether participants' speech-language abilities, age and gender influence the frequency of their speech disfluencies the following covariates were entered into the generalized linear regression model of each dependent variable (stuttered, non-stuttered and total disfluencies): GFTA standard score, PPVT standard score, EVT standard score, TELD receptive subtest standard score, TELD expressive subtest standard score, age, and gender. The model tested main effects of talker group and gender, the talker group x gender interaction and main effects of all covariates.

3.3.1. Stuttered disfluencies—As might be expected based on group classification criteria, analyses indicated a significant main effect of group (Wald $\chi^2 = 912.27$, df = 1, $p < .0001$) for stuttered disfluencies, with CWS exhibiting more stuttered disfluencies than CWNS. There was no interaction between group and gender. None of the covariates in the model were significant, failing to support hypothesis 3 for stuttered disfluencies.

The beta coefficients (i.e., estimates of effect size) for the group main effect in the regression model were as follows (with CWS boys, who produced the most stuttered disfluencies, as the reference): $\beta = -2.045$ for CWNS girls and $\beta = -1.973$ for CWNS boys, and $\beta = -0.100$ for CWS girls. Negative beta weights indicate that, relative to CWS boys, all other groups produced fewer stuttered disfluencies.

3.3.2. Non-stuttered disfluencies—In general, although not included as a part of the CWS versus CWNS classification criteria, results of the analysis for non-stuttered disfluencies indicated four significant main effects, one for group (Wald $\chi^2 = 12.26$, df = 1, $p < .0001$), one for gender (Wald $\chi^2 = 6.05$, df = 1, $p = .014$), one for EVT standard score (Wald $\chi^2 = 6.66$, df = 1, $p = .010$) and one for age (Wald $\chi^2 = 4.92$, df = 1, $p = .027$). There was no significant interaction between group and gender. These findings support hypotheses 2 and 3 for non-stuttered disfluencies. No other covariates (GFTA, PPVT, TELD receptive and expressive subtests standard scores) were significant in the model.

Specifically, regardless of gender, the group effect indicated that CWS produced more non-stuttered disfluencies than CWNS. Further, regardless of talker group, boys produced more non-stuttered disfluencies than girls. The beta coefficients for the group and gender main effects are as follows (with CWS boys, who produced the most non-stuttered disfluencies, as the reference): $\beta = -0.385$ for CWNS girls, $\beta = -0.283$ for CWNS boys and $\beta = -0.212$ for CWS girls.

Although statistically significant, it should be noted that the mean group difference in the frequency of non-stuttered disfluencies was only .86% (CWNS on average produced 3.05% non-stuttered disfluencies and CWS produced 3.91%). Thus, this mean difference appears to be of minimal clinical significance. However, the fact that the difference between the two groups was statistically significant warranted further analyses. Specifically, we attempted to determine whether a cut-off score, rather than group means, might have a stronger clinical

significance and be useful to augment, but not supplant, the presently used 3% stuttered disfluency classification criterion (see Section 3.4 for the details of these analyses).

Further, regardless of talker group, children with lower EVT standard scores produced more non-stuttered disfluencies than children with higher EVT scores; moreover, older children produced more non-stuttered disfluencies. The beta weight for the main effect of EVT standard score was $\beta = -0.008$ indicating that EVT standard score may have minimal clinical significance for the frequency of non-stuttered disfluencies. Similarly, the beta weight for the main effect of age was $\beta = .008$, also suggesting that the effect of age may be of minimal clinical significance.

3.3.3. Total disfluencies (i.e., stuttered + non-stuttered disfluencies)—As might be expected based on group classification criteria, there was a significant main effect of group (Wald $\chi^2 = 499.68$, df = 1, $p < .0001$) with CWS producing more total disfluencies than CWNS. There was also a significant main effect of gender (Wald $\chi^2 = 4.45$, df = 1, $p = .035$) and GFTA standard score (Wald $\chi^2 = 4.33$, df = 1, $p = .037$). The beta weights for the main effects of group and gender are as follows (with CWS boys, who produced the most total disfluencies, as the reference): $\beta = -1.198$ for CWNS girls, $\beta = -1.104$ for CWNS boys, and $\beta = -0.110$ for CWS girls. The beta weight for the main effect of GFTA standard score was $\beta = .004$ indicating minimal clinical significance of this variable for the total disfluency frequency, hence warranting no further consideration for the purposes of this study. There was no significant interaction between group and gender. No other covariates (PPVT, EVT, TELD receptive and expressive subtests standard scores or age) were significant in the model. Thus, hypothesis 2 and 3 were supported for total disfluencies. Mean between-group and between-gender differences on the three dependent variables are depicted in Figs. 2 and 3, respectively.

3.4. Follow-up analysis: frequency of non-stuttered and total disfluencies as a predictor of talker group classification

To determine whether children could be classified as CWS or CWNS based on the frequency of their non-stuttered disfluencies and their total disfluencies, two logistic regression models, models not impacted by non-normality of distribution, were fitted to the data. The talker group classification based on percent stuttered disfluencies and scores on SSI-3 was the dependent variable in both analyses.

Findings indicated that the number of *non-stuttered* disfluencies is a significant predictor of group membership (Wald $\chi^2 = 16.69$, df = 1, $p < .0001$; $\beta = .061$), with 71% of CWNS and 41% of CWS correctly classified based on the frequency of non-stuttered disfluencies. Moreover, the number of *total* disfluencies was a significant predictor of group membership (Wald $\chi^2 = 111.99$, df = 1, $p < .0001$; $\beta = .263$), with 91.4% of CWNS and 85% of CWS correctly classified based on the frequency of total disfluencies. Both classification tables are in Tables 4 and 5.

3.5. Follow-up analysis: sensitivity and specificity analysis for non-stuttered and total disfluencies

Based on the finding that the number of non-stuttered disfluencies, as well as the number of total disfluencies, were significant predictors of group membership, a receiver operating characteristic curve (ROC) analysis (see Swets, 1992) was used to select the optimal threshold to be used in talker group classification.[6] This analysis was necessary because sensitivity (true positives) and specificity (true negatives) depend on the chosen cut point.

Although the threshold criterion to classify children as stuttering or normally fluent based on frequency of their disfluencies potentially can be set anywhere along the continuous scale of disfluency frequency, in the present study we selected what is known to be a "strict approach" (Swets, 1992). This approach sets a criterion that yields few false positive classifications (i.e., classifying someone as a CWS when, in fact, he or she is normally fluent). Hence we set our criterion to yield false positive classifications at .05 (similar to hypothesis testing procedures in which alpha is set to .05 or less).

Thus, with specificity of .95, the 7% non-stuttered disfluencies criterion was identified as a threshold for CWS classification. The area under the ROC curve, a measure of strength of predictive capacity of the model over all cut points, for non-stuttered disfluencies was .61. This was better than chance (.50) but far from perfect (1.00). This indicated that the model fits moderately well and has fair discriminatory ability (Petrie & Sabin, 2009), however the sensitivity of this model is relatively low (11%). The sensitivity–specificity analysis for non-stuttered disfluencies is presented in Table 6. This table shows the trade-off between sensitivity and specificity. For example, a cut point of 10% non-stuttered disfluencies produces high specificity (99%) but low sensitivity (2%).

The same approach was adopted for sensitivity–specificity analysis for number of total disfluencies. With the specificity criterion set at .95 (yielding false positive classifications on the order of .05), 8% total disfluencies was identified as a threshold for CWS classification. The area under the ROC curve for total disfluencies was .958, suggesting a very strong discriminatory ability at almost any cutting score. The sensitivity–specificity analysis for total disfluencies is presented in Table 7.

3.6. Hypotheses 4: stuttered disfluencies and parental concern

To determine the association between parental concern about their child's stuttering and examiner's judgment of stuttering based on frequency of stuttered disfluencies, we employed a logistic regression analysis. A talker group classification based solely on expressed parental concern about the child's stuttering was the dependent variable in this analysis. Frequency of stuttered disfluencies was the independent variable.

The sample for this analysis included the same 472 children reported above. Parents of 254 children expressed concerns about their child's stuttering (184 boys, 70 girls, *M(age)* = 48

---

[6]ROC curve plots the sensitivity of the model against (1 – the specificity) of the model for different threshold of the predicted probability. Sensitivity is defined as the percent of cases correctly identified to have a condition/disease, and specificity – as the percent of cases correctly identified to be "condition-free"/healthy.

months), and parents of 218 children expressed no concerns about stuttering (105 boys, 113 girls, *M(age)* = 50 months). Children whose caregivers expressed concerns about stuttering exhibited an average of 8.11% of stuttered (range: .33–33.67%) and 3.74% of non-stuttered disfluencies (range: 0–12.33%) in their conversational speech. Children whose caregivers did not express concern about stuttering exhibited an average of 1.52% (range: 0–10.67%) of stuttered and 3.15% (range: 0–11%) of non-stuttered disfluencies in their speech.

Logistic regression model fitted to the data indicated that the number of stuttered disfluencies is a significant predictor of parental concern about stuttering (Wald χ^2 = 94.45, df = 1, p < .0001; β = .262), with 90.8% of children whose parents are not concerned about stuttering and 82.3% of children whose parents are concerned correctly classified based on the frequency of stuttered disfluencies. The classification table is presented in Table 8.

Using parental concern as a means for talker-group classification, the present authors sought to determine the sensitivity and specificity of the 3% stuttered disfluencies criterion (e.g., Conture, 2001; Yairi & Ambrose, 2005). In other words, is the 3% criterion a reasonable means for talker-group classification when parental concern is the "gold standard?" The area under the ROC curve, a measure of strength of predictive capacity of the model over all cut points, for stuttered disfluencies was .91. This indicated that the model has good discriminatory ability. Using 3% stuttered disfluencies as a cut-off score for talker-group classification resulted in sensitivity of .80 (true positive classifications) and specificity of .92 (yielding false positive classifications on the order of .08), suggesting that the 3% criterion has a strong and clinically meaningful association with parental concern. The sensitivity–specificity analysis for stuttered disfluencies is presented in Table 9.

4. Discussion

The present study resulted in four main findings: first, frequency distributions of three common disfluency types (stuttered, non-stuttered and total disfluencies) were non-normal. They followed a negative binomial distribution, a Poisson-like count with larger dispersion than true Poisson. Second, there was a significant difference between preschool-age CWS and CWNS in frequency of stuttered as well as non-stuttered disfluencies. Furthermore, the number of non-stuttered and total disfluencies were significant predictors for talker group classification. Third, for both talker groups, expressive vocabulary (as measured by the EVT) and age were associated with the frequency of non-stuttered disfluencies. Moreover, gender was associated with the frequency of non-stuttered and total disfluencies in both talker groups. Fourth, parental concern about children's stuttering was significantly associated with frequency of children's stuttered disfluencies. These findings will be discussed immediately below.

Number of disfluencies is not normally distributed—Present findings that frequency distributions of speech disfluencies were non-normal are consistent with earlier observations (Davis, 1939; Johnson et al., 1963; Jones et al., 2006). The distributions of total, stuttered and non-stuttered disfluencies found in the present study conformed best to a negative binomial distribution. This type of distribution can be characteristic of variables that represent count (i.e., discrete) data. This distribution is often used to model the

occurrence of relatively rare events, such as, in our case, the number of disfluencies children produce during a conversational sample. As applied to the present speech disfluency data set, negative binomial distribution of frequency of disfluencies signifies that there are more cases of mild stuttering among CWS and fewer cases of severe stuttering.

From a data analytic standpoint, the fact that disfluency count data is not-normally distributed suggests that traditional inferential, parametric statistical methods such as ANOVA or ordinary least squares regression are inappropriate for these data. In such cases the mean and variance may not be good descriptors of the central tendency, leading to a potential increase of type 1 error. Going forward, when empirically studying the speech disfluenicies of children who do and do not stutter, it may be more appropriate to employ models that make assumptions that the data actually meet. Generalized linear models (GLM), as used in the present study, allow a choice among several distributions in which the response or dependent variable can have a non-normal distribution (Nelder & Wedderburn, 1972).

Table 10 presents frequency of disfluencies found in the present study and in previous studies of children who do and do not stutter. Although tempting, it is not possible to make absolute comparisons between the present dataset and other studies that also collected comparably large samples (e.g., Johnson et al., 1959; Yairi & Ambrose, 2005; Yaruss, LaSalle, et al., 1998; Yaruss, Max, et al., 1998). This is due to the fact that some of these studies (e.g., Johnson et al., 1959) included children older than the age range of the present study and/or did not report a typically fluent comparison group (e.g., Yaruss, LaSalle, et al., 1998; Yaruss, Max, et al., 1998) and other studies employed a syllable-level measure of frequency (Ambrose & Yairi, 1999; Yairi & Ambrose, 2005).[7]

Thus, even though the present findings of mean values of 1.2 stuttered disfluencies per 100 words for CWNS and 9.2 for CWS is close to the mean values of 1.88 for CWNS and 11.5 for CWS reported by Johnson et al. (1959) and the mean value of 10.67 for CWS reported by Yaruss, LaSalle, et al. (1998) readers should be aware that differences in age range of participants and/or measurement methodology render absolute comparisons problematic. Likewise, there are challenges with making direct comparisons between the present relatively large dataset and other smaller datasets, since larger sample sizes generally lead to increased precision when estimating unknown parameters such as the population mean (Loeve, 1977).

Stuttered and non-stuttered disfluencies—Our second finding that preschool-age CWS produce significantly more stuttered and non-stuttered disfluencies than CWNS corroborates findings from previous studies (Ambrose & Yairi, 1999; Johnson et al., 1959; Yairi & Ambrose, 2005). Whereas the frequency of stuttered disfluencies has been commonly used as a talker-group classification criterion, our data suggest that non-stuttered disfluencies could also be employed to augment decisions about talker group classification based on stuttered disfluencies. The finding that preschool-age CWS produce significantly

---

[7]Present authors recognize that syllable-level measures of stuttering can be converted to word-level measures of stuttering and vice versa (Yaruss, 2001). However, this issue goes beyond the purpose and scope of the present study.

more non-stuttered disfluencies than CWNS and that the number of non-stuttered disfluencies was a significant predictor for talker group classification provides empirical support for the notion that total number of disfluencies may be another augmentative measure useful for distinguishing between children who do and do not stutter (Adams, 1977).

One seemingly apparent assumption, whether children are classified according to parental report (e.g., Boey et al., 2007; Johnson et al., 1959) or objective criteria (e.g., Pellowski & Conture, 2002), is that the speech disfluencies exhibited by CWS versus those of CWNS are more dimensional (i.e., continuous) than categorical (i.e., non-continuous) in nature. Our data suggests that both talker groups produce instances of stuttered disfluencies as well as speech disfluencies not classified as stuttering. Thus, the disfluency distributions for the two talker groups overlap to some degree (something earlier discussed and/or recognized by Johnson et al., 1963). This, of course, does not mean that the two groups are identical. Neither does this overlook the fact that some individuals close to the between-group classification criterion will be challenging to classify. However, clinicians and researchers alike must make decisions about who does and who does not stutter when attempting to empirically study or clinically treat such children.

One attempt to inform this decision-making process or minimize behavioral overlap between the two talker groups is the establishment of a priori criteria for talker group classification (taking into consideration empirical evidence, as well as parental, caregiver and/or professional perceptions). The present finding that the number of non-stuttered disfluencies significantly predicted talker group classification support the use of that variable as an adjunct to (but certainly not replacement for) the 3% stuttered disfluencies criterion for talker group classification.

It should be noted, however, that while minimizing one type of error (e.g., false negatives) this practice may increase the chances of false positives (see Conture, 2001, Fig. 1.1, for further discussion of the issue of false positives and false negatives when classifying children as CWS vs. CWNS). At present, it seems safe to say that there are no absolute, error-free demarcations that perfectly (i.e., 100% of the time) separate the two talker groups. However, as movement toward a more data-based, less error-prone procedure, present findings that 7% non-stuttered disfluency criterion is highly specific and should result in greater accuracy in talker-group classification and help augment the accuracy of the existing 3% stuttered disfluency criterion when employed conjointly.

Influence of expressive vocabulary on preschoolers' non-stuttered disfluencies—In partial support of the third hypothesis, expressive vocabulary size, as measured by the EVT, was shown to be associated with the frequency of non-stuttered disfluencies in both talker groups. For both talker groups, children who exhibited lower expressive vocabulary scores, produced more non-stuttered disfluencies. This finding corroborates an existing body of research suggesting that children's language skill and non-stuttered or "normal" disfluencies are related (Boscolo et al., 2002; Wagovich, Hall, & Clifford, 2009; Westby, 1979). In fact, literature on sentence formulation in adults and the literature on fluency and language interactions in children who stutter and children with

normal fluency (Bernstein Ratner, 1997; Boscolo et al., 2002; Masterson & Kamhi, 1991; Richels et al., 2010; Yaruss et al., 1999; Zackheim & Conture, 2003) suggest that language formulation difficulties or task variations contribute to fluency breakdown.

Perhaps, children with a smaller expressive vocabulary may experience – during typical conversational discourse – more word-finding difficulties, leading to a larger number of non-stuttered disfluencies. It should be kept in mind, however, that the relation between EVT standard score and frequency of non-stuttered disfluencies, although statistically significant, is very subtle ($\beta = -0.008$) and thus may have minimal clinical significance. Indeed, standardized tests may be less than sensitive to the dynamic, rapid and subtle conversational interaction between children's speech disfluencies and concurrent syntactic, lexical and phonological/articulatory processes. Thus, a more comprehensive understanding of this interaction, we suggest, most likely must await further empirical study.

Influence of age on preschoolers' speech disfluencies—In partial support of the third hypothesis, we found that age was associated with the frequency of non-stuttered disfluencies, such that older preschool-age children produced more non-stuttered disfluencies. This association is consistent with Ambrose and Yairi's (1999) finding of a non-significant trend for increase of non-stuttered disfluencies with age in their sample of preschool-age children who do and do not stutter. Of course, children's preschool years (2–5 years of age) represents a time of rapid development of speech and language. Indeed, the present finding that older preschool-age children produced more normal disfluencies seems to suggest that the quantitative and qualitative changes in language that happen during this age may be associated with an increase in non-stuttered disfluencies. However, similarly to the EVT association with disfluencies, the association between age and frequency of non-stuttered disfluencies was very subtle ($\beta = .008$) and thus may have minimal clinical significance.

Influence of gender on preschoolers' non-stuttered and total disfluencies—In partial support of the third hypothesis, we found that gender was associated with the frequency of non-stuttered and total disfluencies in both groups. Higher prevalence of stuttering in boys has been well documented, especially in cases of persistent developmental stuttering (Bloodstein & Bernstein Ratner, 2008; Yairi & Ambrose, 2005). However, few studies have assessed whether gender is related to normal (non-stuttered) disfluency or the extent to which boys and girls differ in frequency of non-stuttered disfluencies.

Our results are consistent with the notion that regardless of talker group (i.e., CWS or CWNS) boys produce more non-stuttered disfluencies (and consequently more total disfluencies) than girls, a finding similar to previous studies (Davis, 1939; Johnson et al., 1959). One possible interpretation of this finding is that boys' speech and language abilities, which are thought to be less robust and/or develop at a slower rate than those of girls of the same age, predispose them to more frequent breakdowns in fluency.

Parental concern for stuttering and stuttered disfluencies—In support of our fourth hypothesis we found a strong association between examiners' judgment of frequency of stuttered disfluencies and parents' expressed concern about stuttering, similar to findings

of an empirical study of a large sample of children who speak Dutch (Boey et al., 2007). This was an expected finding since most parents of children who are brought to a research setting where stuttering is being studied are almost, by definition, interested in and/or have concerns about their child's fluency. Whether the concern stems from parental perceptions of their child's speech disfluency, the fact that the parent or other family member stutters, relatives, teachers or health care professions suggesting that the child is stuttering, or some combination of two or more of these conditions, parental concern about their child's "stuttering" exists for many parents assessed clinically as well as for the purposes of research. Thus, the precise role of parental concern relative to the diagnosis and treatment of childhood stuttering remains a viable topic for empirical study.

At least knowing that these two variables – examiners' judgment of frequency of stuttered disfluencies and parental concern about a child's stuttering – were significantly associated provides empirical evidence for the use of frequency of stuttered disfluencies as an objective, robust criterion to classify children as those who do and do not stutter. However, as with most "objective" talker group classification criteria (e.g., the 3% stuttered disfluencies criterion), some children will present in both clinical as well as research settings to be on the margins or cusp of this criterion, making them challenging to classify.

4.1. Caveats

One salient caveat pertaining to the current study is that our paradigm did not allow for an exhaustive assessment of speech and language ability. For example, articulation abilities, assessed with GFTA, were tested only at a single-word level. Moreover, language competence was largely evaluated by expressive and receptive vocabulary tests (PPVT and EVT), and only one test of overall language ability was employed (TELD). Replication of the present study with more conversational-based or direct measures of language, for example, mean length of utterance (MLU), is recommended to better understand how the present measures of (non)stuttered disfluencies are associated with actual speech-language planning and production.

Also, the disfluency counts were based only on the first 300 words of conversational speech. Clinical knowledge about stuttering shows that stuttering waxes and wanes not only from day to day, but also frequency of disfluency may increase with the size of language sample (Sawyer & Yairi, 2006). Moreover, there is a possibility that the real-time disfluency counting procedure employed in the current study may be more prone to errors than an off-line disfluency count from recorded speech (see Yaruss, 1998a,b). Whether these differences make a significant difference in terms of distinguishing children who do stutter from those children who do not stutter would appear to be an open, empirical question. It should be noted, however, that Yaruss and colleagues (1998) assessed reliability between the on-line and off-line transcript-based methods of disfluency counts. They reported a strong positive correlation of $r = .88$ ($p < .001$) between on-line and transcript-based count of non-stuttered disfluencies and no significant mean differences between the two counts. Similarly, for stuttered disfluencies Yaruss et al. reported a strong positive correlation of $r = .94$ ($p < .001$) and no significant differences in mean stuttered disfluencies between the online and the transcript-based method of count.

5. Conclusion

Although it was not the purpose of the present study to empirically test extant theories of childhood stuttering, our study did assess four salient issues germane to classification and measurement of childhood stuttering in preschool-age children. Among those issues are non-normality of distributions of speech disfluencies; inclusion of frequency of non-stuttered disfluencies as an additional criterion for talker-group classification to augment the 3% stuttered disfluencies criterion in difficult-to-classify cases, and gender effects on overall fluency of preschool-age children regardless of their talker-group classification.

Results of this study provide additional perspective on findings reported by Johnson et al. (1959) as well as others (e.g., Jones et al., 2006). Whereas our findings do not provide an absolute, error-free means for classifying preschoolers who stutter (something seldom possible, if ever, for classification schemas for human disorders), present findings do move us toward a more comprehensive understanding of salient behaviors and characteristics of preschool-age children's stuttering as well as the empirical means to measure same.

Acknowledgments

References

Adams MR. A clinical strategy for differentiating the normally nonfluent child and the incipient stutterer. Journal of Fluency Disorders. 1977; 2:141–148.

Ambrose N, Yairi E. Normative disfluency data for early childhood stuttering. Journal of Speech, Language and Hearing Research. 1999; 42:895–909.

Anderson JD, Conture EG. Language abilities of children who stutter: A preliminary study. Journal of Fluency Disorders. 2001; 25(4):283–304.

Arnold H, Conture E, Walden T, Key A. Emotional reactivity, emotional regulation and childhood stuttering: A behavioral, electrophysiological study. Journal of Communication Disorders. 2011; 44:276–293. [PubMed: 21276977]

Bernstein, Ratner N. Stuttering: A psycholinguistic perspective. In: Curlee, R.; Siegel, G., editors. Nature and treatment of stuttering: New directions. 2. Boston: Allyn & Bacon; 1997. p. 99-127.

Bloodstein, O.; Bernstein, Ratner N. A handbook on stuttering. Clifton Park, NY: Thompson/Delmar; 2008.

Blood G, Ridenour V, Qualls C, Hammer C. Co-occurring disorders in children who stutter. Journal of Communication Disorders. 2003; 36:427–448. [PubMed: 12967738]

Boey R, Wuyts F, Van de Heyning P, Bodt M, Heylen L. Characteristics of stuttering-like disfluencies in Dutch-speaking children. Journal of Fluency Disorders. 2007; 32:310–329. [PubMed: 17963939]

Boscolo B, Bernstein Ratner N, Rescorla L. Fluency characteristics of children with a history of specific expressive language impairment (SLI-E). American Journal of Speech-Language Pathology. 2002; 11:41–49.

Byrd C, Conture E, Ohde R. Incremental and holistic phonological priming of young children who stutter. American Journal of Speech-Language Pathology. 2007; 16:43–53. [PubMed: 17329674]

Carlo EJ, Watson JB. Disfluencies of 3- and 5-year old Spanish-speaking children. Journal of Fluency Disorders. 2003; 28:37–53. [PubMed: 12706912]

Choi D, Conture EG, Walden TA, Lambert W, Tumanova V. Behavioral inhibition and childhood stuttering. Journal of Fluency Disorders. 2013; 38:171–183. [PubMed: 23773669]

Clark CE, Conture E, Frankel C, Walden T. Communicative and psychological dimensions of the KiddyCAT. Journal of Communication Disorders. 2012; 45:223–234. [PubMed: 22333753]

Clark CE, Conture EG, Walden TA, Lambert WE. Speech sound articulation abilities of preschool-age children who stutter. Journal of Fluency Disorders. 2013; 38(4):325–341. [PubMed: 24331241]

Conture, E. Stuttering: Its nature, diagnosis, and treatment. Needham Heights, MA: Allyn & Bacon; 2001.

Davis DM. The relation of repetitions in the speech of young children to certain measures of language maturity and situational factors: Part I. Journal of Speech Disorders. 1939; 4:303–318.

DeJoy D, Gregory H. The relationship between age and frequency of disfluency in preschool children. Journal of Fluency Disorders. 1985; 10:107–122.

Dunn, L.; Dunn, L. Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4). Circle Pines, MN: American Guidance Service Inc; 2007.

Einarsdottir J, Ingham R. Have the disfluency-type measures contributed to the understanding and treatment of developmental stuttering? American Journal of Speech-Language Pathology. 2005; 14:260–273. [PubMed: 16396610]

Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological Bulletin. 1995; 118(3):392–404. [PubMed: 7501743]

Goldman, R.; Fristoe, M. Goldman–Fristoe Test of Articulation-Second Edition (GFTA-2). Circle Pines, MN: American Guidance Services Inc; 2000.

Hall NE, Yamashita TS, Aram DM. Relationship between language and fluency in children with developmental language disorders. Journal of Speech and Hearing Research. 1993; 36:568–579. [PubMed: 7687313]

Hardin, JW.; Hilbe, J. Generalized estimating equations. Boca Raton, FL: Chapman & Hall/CRC; 2003.

Hollingshead, A. Four factor index of social status. New Haven, CT: Yale University; 1975. unpublished manuscript

Hresko, W.; Reid, D.; Hammill, D. Test of Early Language Development-3. Austin, TX: PRO-ED; 1999.

Johnson, W.; Boehmler, R.; Dahlstrom, W.; Darley, F.; Goodstein, L.; Kools, J., et al. The onset of stuttering. Minneapolis: University of Minnesota Press; 1959.

Johnson, W.; Darley, F.; Spriestersbach, D. Diagnostic methods in speech pathology. New York: Harper & Row; 1963.

Johnson K, Conture E, Walden T. Efficacy of attention regulation in preschool-age children who stutter: A preliminary investigation. Journal of Communication Disorders. 2012; 45:263–278. [PubMed: 22560538]

Johnson KN, Walden TA, Conture EG, Karrass J. Spontaneous regulation of emotions in preschool children who stutter: Preliminary findings. Journal of Speech, Language and Hearing Research. 2010; 53(6):1478–1495.

Jones M, Onslow M, Packman A, Gebski V. Guidelines for statistical analysis of percentage of syllables stuttered data. Journal of Speech Language and Hearing Research. 2006; 49:867–878.

Logan KJ. Language and fluency characteristics of preschoolers' multiple-utterance conversational turns. Journal of Speech, Language and Hearing Research. 2003; 46(1):178–189.

Logan KJ, LaSalle LR. Grammatical characteristics of children's conversational utterances that contain disfluency clusters. Journal of Speech Language and Hearing Research. 1999; 42:80–92.

Loeve, M. Probability theory. 4. Vol. I. New York: Springer-Verlag; 1977.

Martins VO, Andrade CRF. Speech fluency developmental profile in Brazilian Portuguese speakers. Pró-Fono Revista de Atualização Científica. 2008; 20:7–12. [PubMed: 18408857]

Masterson JJ, Kamhi AG. The effects of sampling conditions on sentence production in normal, reading-disabled, and language-learning-disabled children. Journal of Speech, Language and Hearing Research. 1991; 34(3):549–558.

McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods. 1996; 1(1):30–46.

Natke U, Sandrieser P, Pietrowsky R, Kalveram KT. Disfluency data of German preschool children who stutter and comparison children. Journal of Fluency Disorders. 2006; 31:165–176. [PubMed: 16837037]

Nelder J, Wedderburn R. Generalized linear models. Journal of the Royal Statistical Society. 1972; 135:370–384.

Nippold M. Stuttering and language abilities in children: Questioning the connection. American Journal of Speech-Language Pathology. 2012; 21:183–196. [PubMed: 22442282]

Ntourou K, Conture E, Lipsey M. Language and childhood stuttering: A meta-analytical approach. American Journal of Speech-Language Pathology. 2011; 20:163–179. [PubMed: 21478281]

Ntourou K, Conture EG, Walden TA. Emotional reactivity and regulation in preschool-age children who stutter. Journal of Fluency Disorders. 2013; 38:260–274. [PubMed: 24238388]

Pellowski M, Conture E. Characteristics of speech disfluency and stuttering behaviors in 3- and 4-year-old children. Journal of Speech, Language and Hearing Research. 2002; 45:20–34.

Petrie, A.; Sabin, C. Medical statistics at a glance. Malden, MA: Blackwell Science Ltd; 2009.

Richels C, Buhr T, Conture E, Ntourou K. Utterance complexity and stuttering on function words of preschool-age children who stutter. Journal of Fluency Disorders. 2010; 35(3):314–331. [PubMed: 20831974]

Richels, C.; Conture, E. Diagnostic data as predictor of treatment outcome with young children who stutter. In: Guitar, B.; McCauley, R., editors. Treatment of stuttering: Conventional and controversial interventions. Baltimore, MD: Lippincott Williams & Wilkins; 2010. p. 18-55.

Riley, G. Stuttering Severity Instrument of Children and Adults. 3. Tigard, OR: CC Publications; 1994.

Sawyer J, Yairi E. The effect of sample size on the assessment of stuttering severity. American Journal of Speech and Language Pathology. 2006; 15:36–44.

Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965; 52(3–4):591–611.

Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin. 1979; 86(2):420–428. [PubMed: 18839484]

Swets JA. The science of choosing the right decision threshold in high-stakes diagnostics. American Psychologist. 1992; 47:522–532. [PubMed: 1595983]

Vanryckeghem, M.; Brutten, G. The KiddyCAT: Communication attitude test for preschool and kindergarten children who stutter. San Diego, CA: Plural Publishing; 2007.

Wagovich SA, Hall NE, Clifford BA. Speech disruptions in relation to language growth in children who stutter: An exploratory study. Journal of Fluency Disorders. 2009; 34:242–256. [PubMed: 20113769]

Walden TA, Frankel CB, Buhr AP, Johnson KN, Conture EG, Karrass JM. Dual diathesis-stressor model of emotional and linguistic contributions to developmental stuttering. Journal of Abnormal Child Psychology. 2012; 27:145–154.

Watkins RV, Yairi E. Language production abilities of children whose stuttering persisted or recovered. Journal of Speech, Language and Hearing Research. 1997; 40:385–399.

Westby CE. Language performance of stuttering and nonstuttering children. Journal of Communication Disorders. 1979; 12(2):133–145. [PubMed: 429605]

Williams, KT. Expressive Vocabulary Test. 2. Circle Pines, MN: American Guidance Services Inc; 2007. (EVT-2)

Yairi, E.; Ambrose, NG. Early childhood stuttering: For clinicians by clinicians. Austin, TX: PRO-ED; 2005.

Yairi E, Ambrose N, Paden EP, Throneburg RN. Predictive factors of persistence and recovery: Pathways of childhood stuttering. Journal of Communication Disorders. 1996; 29:51–77. [PubMed: 8722529]

Yairi E, Clifton NF Jr. Disfluent speech behavior of preschool children, high school seniors, and geriatric persons. Journal of Speech and Hearing Research. 1972; 15:714–719. [PubMed: 4652390]

Yairi E, Lewis B. Disfluencies at the onset of stuttering. Journal of Speech, Language and Hearing Research. 1984; 27:145–154.

Yaruss J. Clinical implications of situational variability in preschool children who stutter. Journal of Fluency Disorders. 1998a; 22:187–203.

Yaruss J. Real-time analysis of speech fluency: Procedures and reliability training. American Journal of Speech-Language Pathology. 1998b; 7:25–37.

Yaruss JS. Converting between word and syllable counts in children's conversational speech samples. Journal of Fluency Disorders. 2001; 25(4):305–316.

Yaruss J, LaSalle L, Conture E. Evaluating stuttering in young children. American Journal of Speech-Language Pathology. 1998; 7:62–76.

Yaruss J, Max M, Newman R, Campbell J. Comparing real-time and transcript-based techniques for measuring stuttering. Journal of Fluency Disorders. 1998; 23:137–151.

Yaruss J, Newman RM, Flora T. Language and disfluency in nonstuttering children's conversational speech. Journal of Fluency Disorders. 1999; 24:185–207.

Yoder, PJ.; Symons, FJ. Observational measurement of behavior. New York, NY: Springer Publishing Company; 2010.

Zackheim CT, Conture EG. Childhood stuttering and speech disfluencies in relation to children's mean length of utterance: A preliminary study. Journal of Fluency Disorders. 2003; 28(2):115–142. [PubMed: 12809748]

Appendix A. Continuing education

CEU questions

1. Non-stuttered disfluencies are

    a. More frequently produced by children who stutter.

    b. More frequently produced by normally fluent children.

    c. More frequently produced by boys rather than girls, regardless of talker group.

    d. Both (a) and (c).

2. Which distribution best approximates the distribution of speech disfluencies of preschool-age children in the present study?

    a. Gaussian.

    b. Multinomial.

    c. Negative binomial.

    d. Gamma.

3. Which disfluency type significantly differentiates children who stutter from their normally fluent peers?

    a. Stuttered disfluencies.

    b. Non-stuttered disfluencies.

    c. Total disfluencies.

    d. All of the above.

4. Older children in the present study produced more stuttered disfluencies.

True/False

5. With the specificity criterion set at .95, which percent of total disfluencies in conversational speech was identified as a threshold for talker-group classification?

    a. 5%

    b. 7%

    c. 8%

    d. 12%
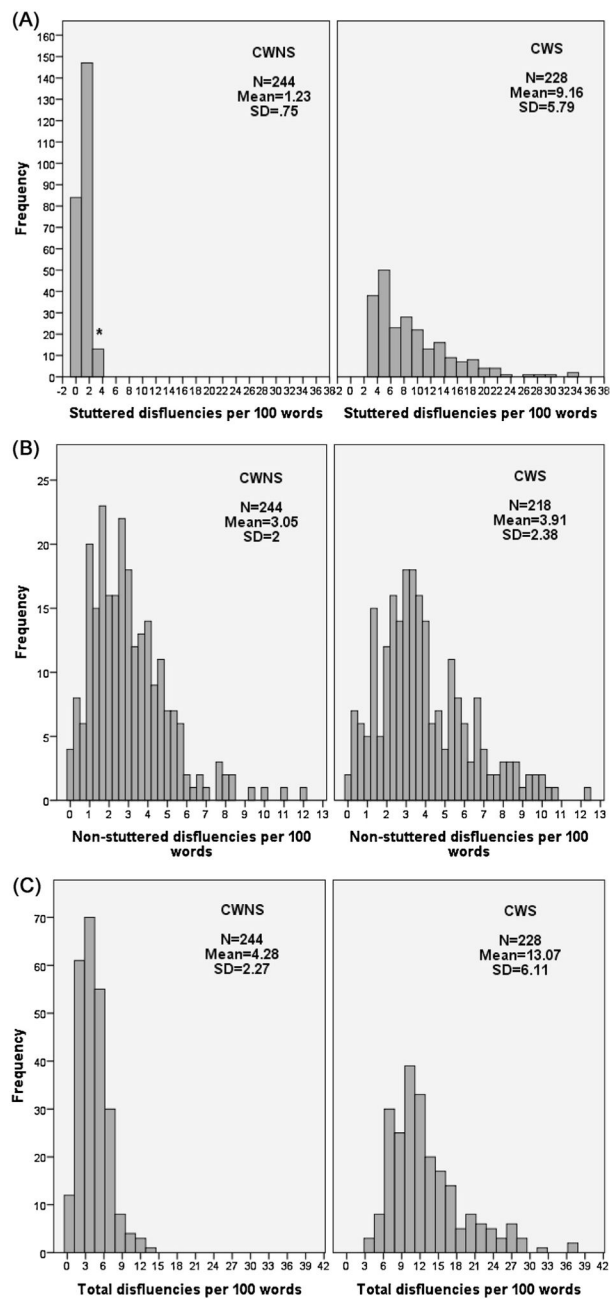
Fig. 1.
Histogram of (A) stuttered disfluencies, (B) non-stuttered disfluencies, and (C) total disfluencies per 100 words of conversational speech for preschool-age children who stutter (CWS; *n* = 228) and children who do not stutter (CWNS; *n* = 244). **Note*: In figure (A) 9 CWNS had 2.6 stuttered disfluencies per 100 words. The graphical software used to make this figure rounded up that value to 3.

Fig. 2.
Mean group differences (brackets = 95% confidence interval, CI) in stuttered, non-stuttered and total disfluencies per 300 words between preschool-age children who stutter (CWS; *n* = 228) and children who do not stutter (CWNS; *n* = 244).

Fig. 3.
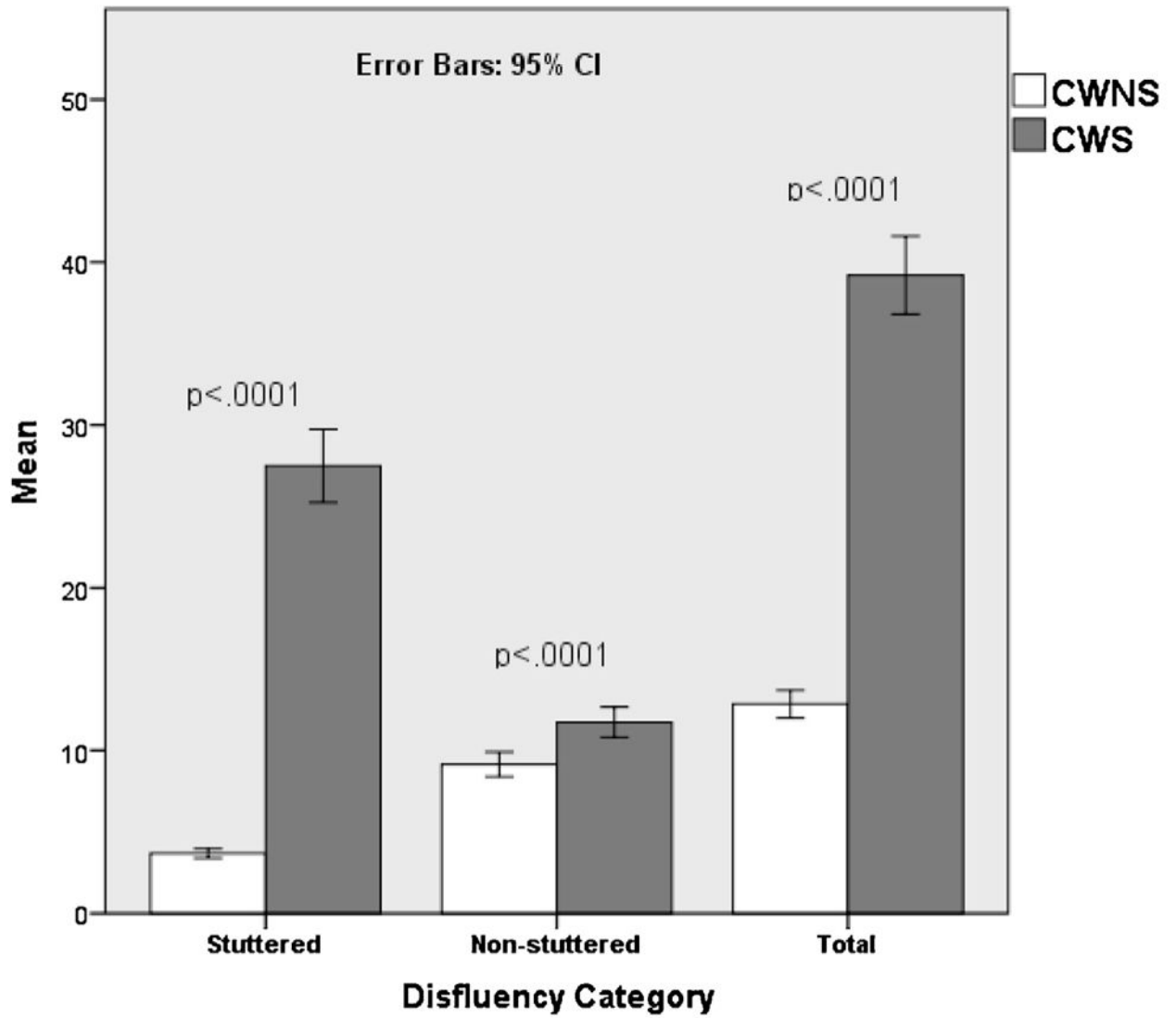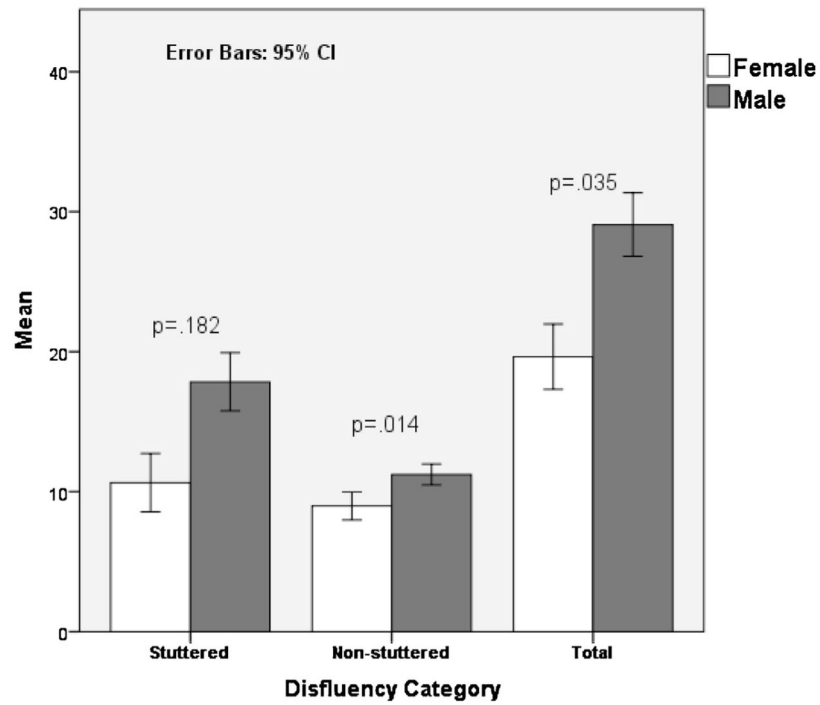Mean gender differences (brackets = 95% confidence interval, CI) in stuttered, non-stuttered and total disfluencies per 300 words for all participants, that is, preschool-age children who stutter (CWS; *n* = 228) plus children who do not stutter (CWNS; *n* = 244).

Table 1

Descriptive information for CWNS (n = 244) and CWS (n = 228) for each of the independent variables.

Independent variable	Talker group	Mean	SD
Age (months)	CWNS	50.47	9.59
	CWS	47.47	8.94
SES family	CWNS	45.23	11.11
	CWS	44.67	11.20
GFTA Standard Score	CWNS	107.58	11.84
	CWS	106.42	11.64
PPVT Standard Score	CWNS	110.77	13.95
	CWS	106.33	14.18
EVT Standard Score	CWNS	113.93	13.03
	CWS	109.70	13.55
TELD Receptive Standard Score	CWNS	115.81	14.14
	CWS	110.39	17.82
TELD Expressive Standard Score	CWNS	106.80	13.42
	CWS	105.34	14.83
TELD Spoken Language Standard Score	CWNS	113.57	14.84
	CWS	108.98	18.43

Table 2

Distribution of stuttered (SD), non-stuttered (NSD), and total (TD) disfluencies.

Type of disfluency	Talker group	Percentiles (data per 100 words)										
		5	10	20	30	40	50	60	70	80	90	95
SDs	CWNS	.00	.33	.67	.67	1	1	1.3	1.7	2	2.3	2.7
	CWS	3	3.7	4.3	5	6	7.7	9	10.7	13.3	17.7	20.4
NSDs	CWNS	.42	1	1.3	1.7	2.3	2.67	3	3.7	4.5	5.3	6.9
	CWS	.67	1.3	2	2.7	3	3.3	4	4.7	5.7	7.2	8.7
TDs	CWNS	1.08	1.8	2.3	3	3.3	4	4.7	5	5.7	7.2	8
	CWS	6.2	6.7	8.3	9.7	10.7	11.7	12.7	14.7	17	22	26.3

Table 3

Descriptive statistics for total disfluencies (TD), stuttering disfluencies (SD), and non-stuttering disfluencies (NSD). Data based on 300 word sample.

Variable		Mean	Median	Variance	STD	SEM[a]	Skewness	Kurtosis
SDs	CWNS	3.69	3	5.01	2.24	.175	.25	-.87
	CWS	27.49	23	309.78	17.37	1.139	1.51	2.80
NSDs	CWNS	9.14	8	35.85	5.99	.367	1.31	2.69
	CWS	11.73	10	51.09	7.15	.545	.86	.51
TDs	CWNS	12.84	12	46.54	6.82	.414	.97	1.46
	CWS	39.21	35	335.62	18.32	1.344	1.33	1.95

[a]Data based on 228 CWS and 244 CWNS; SEM = standard error of the mean.

Table 4

Logistic regression analysis: frequencies and percentages of individuals correctly and incorrectly classified into CWS ($N = 228$) versus CWNS ($N = 244$) groups based on the number of non-stuttered disfluencies.

Actual category	Predicted category		
	CWS	**CWNS**	**% correct**[a]
228 CWS	94	134	41.2
244 CWNS	71	173	70.9

[a]Overall percentage correct is 56.6.

Table 5

Logistic regression analysis: frequencies and percentages of individuals correctly and incorrectly classified into CWS ($N = 228$) versus CWNS ($N = 244$) groups based on the number of total disfluencies.

Actual category	Predicted category		
	CWS	CWNS	% correct[a]
228 CWS	194	34	85.1
244 CWNS	21	223	91.4

[a]Overall percentage correct is 88.3.

Table 6

ROC curve analysis: sensitivity–specificity table for select thresholds of frequency of non-stuttered disfluencies (NSD).

Percent NSD threshold	Sensitivity	Specificity
6	.180	.935
7	.11	.953
8	.077	.971
9	.042	.984
10	.018	.990

Table 7

ROC curve analysis: sensitivity–specificity table for select thresholds of frequency of total disfluencies (TD).

Percent TD threshold	Sensitivity	Specificity
7	.875	.890
8	.818	.945
9	.755	.961
10	.656	.976
11	.551	.984

Table 8

Logistic regression analysis: frequencies and percentages of children whose parents are and are not concerned about stuttering classified into CWS and CWNS groups based on the number of stuttered disfluencies (SD).

Parental concern about stuttering	Predicted category based on SD frequency		
	CWS	CWNS	% correct[a]
254 CWS	209	45	82.3
218 CWNS	20	198	90.8

[a]Overall percentage correct is 86.2.

Table 9

ROC curve analysis: sensitivity–specificity table for select thresholds of frequency of stuttered disfluencies.

Percent SD threshold	Sensitivity	Specificity
0	1	.069
1	.953	.580
3	.803	.918
5	.589	.955
7	.490	.991

Table 10

Comparison of frequencies of stuttered (SD) and non-stuttered (NSD) disfluencies (per 100 words) in different studies of children.

Study	Group	*N*	Mean SD	Mean NSD
Ambrose and Yairi (1999)	CWS	90	10.37[a]	5.0[a]
	CWNS	54	1.33[a]	3.89[a]
Boey et al. (2007)	CWS	693	15.71	n/a
	CWNS	79	.42	n/a
Johnson et al. (1959)	CWS	68M	11.5	6.4
	CWNS	68M	1.88	5.4
Pellowski and Conture (2002)	CWS	36	8.7	2.0
	CWNS	36	1.1	1.5
Yairi and Ambrose (2005)	CWS	103	11.3[a]	5.79[a]
	CWNS	52	1.41[a]	4.48[a]
Yaruss, LaSalle, et al. (1998)	CWS	100	10.67	n/a
	CWNS	n/a	n/a	n/a
Present study	CWS	228	9.2	3.91
	CWNS	244	1.2	3.05

[a]Frequency per 100 syllables.