



## Viewpoint

# Have Disfluency-Type Measures Contributed to the Understanding and Treatment of Developmental Stuttering?

Jóhanna Einarsdóttir

University of Iceland, Reykjavik

Roger J. Ingham

University of California, Santa Barbara

**Purpose:** This article critically reviews evidence to determine whether the use of disfluency typologies, such as *syllable repetitions* or *prolongations*, has assisted the understanding or treatment of developmental stuttering. Consideration is given to whether there is a need for a fundamental shift in the basis for constructing measures of stuttering behavior.

**Method:** The history of using specific types of disfluencies to assess stuttering, including more recent developments such as counts of stuttering-like disfluencies, is reviewed. The focus is on studies that have investigated the validity and reliability of these perceptually based assessment methods.

**Conclusion:** The evidence from use of disfluency-type measures shows that the behavioral difference between stuttering and normally fluent speakers is solely related to the amount of observable stuttering; the differences are only partially realized within disfluency-type measures. Indeed, because disfluency-type measures show poor reliability and conflate stuttered and nonstuttered speech, they have only limited heuristic value for research and provide no obvious benefits for clinicians. At best, they should be regarded as imprecise descriptors of observable stuttering and not a fundamental measure of stuttering. A recommended solution to the problematic history of verbal-based definitions of stuttering behavior is continued development and investigation of exemplar-based definition and measurement.

**Key Words:** stuttering, disfluency types, measurement, validity

---

The way we define a disorder is, as Bloodstein (1995, p. 2) observed, "bound up with the question of the operations we use to measure it." From a scientific point of view, it is necessary that the measurement, and therefore the diagnosis, of a disorder should have consistency if research on that disorder is to make progress. In a disorder such as developmental stuttering, it is also important to be able to evaluate the development

of the disorder and establish the effectiveness of therapies that seek to alleviate its impact. In the past half century, however, different theories of stuttering have greatly influenced the way this disorder is measured. This has resulted in a variety of measurement methods, indeed a conglomeration of methods that may have actually impeded research on stuttering. Consider, for instance, the plethora of measurement protocols that have been proposed to help clinicians distinguish between stuttering and nonstuttering children (Adams, 1977; Ambrose & Yairi, 1999; Campbell & Hill, 1994; Conture, 1997; Curlee, 1980; Gordon & Luper, 1992; J. Ingham & Riley, 1998; R. J. Ingham, 1985; Onslow, 1992; Pindzola & White, 1986; Riley & Riley, 1982). The large number of such protocols may simply reflect the difficulties involved in distinguishing between these children. However, as will be argued, this proliferation could also be due to the historic and philosophically driven reluctance of many researchers and clinicians to simply measure the occurrence of perceived stuttering during speech in order to track the development and variability of this disorder, and to evaluate its treatment. Indeed, this reluctance may have been fostered by a continuing unquestioned tradition of measuring stuttering via one of the various disfluency typologies that originated from those developed by Wendell Johnson in the early 1960s (Johnson, 1961; Johnson et al., 1959).

Perhaps the most prominent disfluency-type metric currently used in stuttering assessment or diagnosis is the number of stuttering-like disfluencies (SLDs) per words or syllables spoken. The SLDs measure was largely pioneered by Yairi and Ambrose (1992); they essentially reclassified a subset of Johnson et al.'s eight categories that were principally designed to describe normal disfluencies. SLDs came to prominence in the 1990s because they were the metric of choice in what became known as the "Illinois Studies." Their use in these studies also prompted considerable controversy (Cordes, 2000b; Cordes & Ingham, 1994, 1995b, 1996; R. J. Ingham & Bothe, 2001; Onslow & Packman, 2001; Packman & Onslow, 1998; Wingate, 2001). Nonetheless, SLDs continue (in one form or another) to be the favored metric within the Illinois Studies. More importantly, this metric (or its variants) appears to be flourishing within other recent studies and books on stuttering children (see Anderson & Conture, 2004; Logan, 2003; Pellowski & Conture, 2002; Trautman, Healey, Brown, Brown, & Jermano, 1999; Trautman, Healey, & Norris, 2001; Zackheim & Conture, 2003;

Zebrowski & Kelly, 2002). For that reason, it is important to ask whether disfluency-type measurement has made a special contribution toward the understanding and treatment of developmental stuttering. More particularly, clinicians need to know whether such measures do make a contribution to assessment and treatment that cannot be achieved *unless* stuttered speech is categorized in terms of types of disfluencies.

In this article we will argue that (a) there is no evidence that the use of disfluency typologies, such as *syllable repetitions* or *prolongations*, to measure and diagnose stuttering behavior has made any significant contribution to the understanding or treatment of this disorder, and (b) more significant contributions will likely emerge from a fundamental shift in the basis for constructing measures of stuttering behavior.

### **Methods of Measuring Stuttering**

Most methods currently used to measure stuttering rely on listener judgments, preferably agreed judgments, of the number of disfluencies that are judged to occur during various speech samples (Bloodstein, 1995; Cordes & Ingham, 1994). These measures typically involve judgments of the type of behaviors within a speech sample and/or global ratings of the speech sample. The focus here will be on measures that have used judgments of certain types of speech behavior.

Over the past 50 years, researchers have tended to favor measuring stuttering by recording instances of one of three superficially related but fundamentally different types of speech events (see Cordes & Ingham, 1994): (a) Johnson et al.'s (1959) different disfluency types, or variations thereof; (b) Wingate's (1964) kernel characteristics of stuttering (a restricted range of disfluency types); or (c) events that are simply perceived to be stuttering events (Martin & Haroldson, 1981).<sup>1</sup> Johnson's Diagnosogenic Theory (Johnson, 1956), first of all, incorporated the very influential claim that there is essentially no difference between the types of disfluencies that typify normally fluent speech and those that characterize the speech of stuttering speakers. Driven by his theory's emphasis on stuttering development, Johnson was especially interested in investigating the types of disfluencies that occur in the speech of young children. Indeed, his theorized explanation for the cause of stuttering relied on identifying such disfluencies; children begin to stutter, he argued, because parents highlight bouts of normal disfluency in their child's speech. The disfluency typology that Johnson and colleagues developed and employed to measure these disfluencies emerged from studies of the speech of college-age male and female stuttering and nonstuttering speakers (Johnson, 1961; Johnson et al., 1959). The initial set of categories constituting this typology was as follows: *interjections*, *sound and syllable repetitions*, *word repetitions*, *phrase*

*repetitions*, *revisions*, *incomplete phrases*, *broken words*, and *prolonged sounds* (Johnson et al., 1959). It is noteworthy that *pause time* was considered for inclusion but was bypassed because judges had difficulty in deciding when a pause was a legitimate part of fluent speech (Johnson et al., 1959). It is also important to take note of Wingate's (1987) observation that these categories were framed by the speech of stuttering speakers and are not commonly used outside of stuttering research to describe disfluent utterances. For that reason, as Wingate noted, these categories offer a very restricted account of speech performance. Ironically, while Johnson contended that stuttered speech and normally disfluent speech were essentially similar, he did introduce the "moment of stuttering" as a unit for quantifying episodes of nonfluency during stuttered speech (Johnson, 1955); arguably, that unit continues to be the principal basis for measuring the frequency of stuttering.

Wingate's (1964) "standard definition of stuttering," the second common conceptualization of stuttering, is a three-part definition based on documented accounts (as of 1964) of the behavior of stuttering speakers. Its principal and distinguishing feature is its description of the so-called "kernel characteristics" of stuttering. The other two parts describe accessory features that appear to be more descriptive than definitive. Wingate's description of the kernel characteristics is actually closely related to Johnson's disfluency descriptors; they are "audible or silent, *repetitions or prolongations* [italics added] in the utterance of short speech elements, namely: sounds, syllables, and words of one syllable" (Wingate, 1964, p. 488). In other words, the focus of Wingate's definition is a subset of Johnson's disfluency types.

Martin and Haroldson (1981) proposed a fundamentally different definition—a more behaviorally based perceptual threshold definition of stuttering. Occasions of stuttering were simply defined as those events that an individual perceives to be "stuttering" because they exceed a perceptual threshold for recognizing stuttering (Martin & Haroldson, 1981). Although they were "individual," these threshold-based judgments were not completely idiosyncratic; they relied ultimately on interjudge agreement for their validity. It is important to recognize that the perceptual threshold definition was anticipated by Bloodstein (1975, p. 10) in his much-quoted observation that stuttering is most likely defined by "whatever is perceived as stuttering by a reliable observer who has relatively good agreement with others."

### **Counting Total Disfluencies**

Yairi (1997, p. 49) has argued that "disfluency counts have been the classic metric of the disorder [stuttering] as well as the dependent variable of interest in both clinical and experimental studies." This metric is indeed "classic," insofar as it has not changed substantially since it was first introduced by Johnson in the late 1950s (Johnson, 1961; Johnson et al., 1959). Typically, a clinician or researcher transcribes a speech sample and then assigns each transcribed disfluency to a presumably exclusive category. It becomes immediately obvious, though, that

<sup>1</sup>It is recognized that other measures of stuttering have been described, for example, Perkins's (1983) "loss of control," Smith and Kelly's (1997) multifactorial assessment, and Howell et al.'s (1997a, 1997b) "lexical disfluencies." However, these measures are rarely used in stuttering research and, seemingly, not at all in treatment.

this methodology requires that no distinction be made between normal disfluencies and stuttering. Most importantly, and most obviously, not one of the categories carries the label “stutter.”<sup>2</sup> As will be seen, variations on Johnson’s method, all without a “stutter” category, continue to be advocated as useful for differentiating between normally disfluent and stuttered speech (Ambrose & Yairi, 1999; Pellowski & Conture, 2002; Yaruss, 1997).

Johnson’s system has been much modified since it first appeared, thereby restricting comparisons among studies that used one of the system’s many variants. For example, Gregory and Hill (1993) and Zebrowski (1991) used a system that included 10 disfluency categories, while Carlo and Watson (2003) used 15 categories. Yairi and colleagues have actually used a number of different categories: Yairi and Ambrose (1992) employed 8, Yairi, Ambrose, and Niermann (1993) used 7, and a 1999 study by Yairi and Ambrose used only 6.

Regardless of which variant of Johnson’s system is used, however, the primary issue continues to be whether any such system can provide reliable and clinically meaningful stuttering measurement data—not disfluency measurement data. Which of these variants can really be justified as the “dependent variable of interest” in stuttering research? The answer should reside in tests to determine whether the frequency per se of disfluencies or disfluency types that characterize an individual’s speech do constitute a valid sign of stuttering (Adams, 1977; Conture, 2001). A number of studies have addressed this issue, but their findings are neither conclusive nor convincing.

A study by Yairi (1982) showed that the frequency of disfluencies in normally fluent 2- to 3-year-old children can vary in unpredictable ways. The children in this study were assessed on three or four occasions at 4-month intervals, and, as expected, all displayed disfluencies on each occasion. Total disfluencies varied greatly, both within and between the children, but the data do not give any indication as to whether a child was stuttering. Consider, for instance, one child (Participant 12) who was reported to stutter during the study. She was classified as normally fluent at her first testing, producing 1.5 disfluencies per 100 words. At her second assessment, both the parents and the investigators agreed she was stuttering; at that time she produced 9.4% disfluencies. During her third assessment, she produced 2.8% disfluencies and was classified as normally fluent, and at her fourth assessment, when she produced 10.0% disfluencies, she was still considered to be normally fluent (Yairi, 1982, Table 3, p. 25). Other children in the study were also very disfluent but were not judged to be stuttering. Participant 15, for example, displayed 25.6% disfluent words during the first assessment, 12.8% during the second, and 12.2% during the third, and was still judged to be a normal speaker. In short,

the total number of disfluencies produced by these children had no obvious relationship to whether they were described as normally fluent or stuttering.

Total disfluency counts among non-English-speaking children have also been investigated. Carlo and Watson (2003), for example, assessed the speech of 32 nonstuttering, monolingual Spanish-speaking children from Puerto Rico, classifying their disfluencies into 15 different categories so as to compare 3-year-old and 5-year-old children. Two participants (female) showed markedly higher frequencies of disfluency than the other children, but according to their teachers and parents they did not stutter. Despite this, the authors used the disfluency data to conclude that it “may be that these girls were not ‘normally fluent’ but actually exhibiting early stuttering” (p. 46). That may be, but the record of the disfluencies produced by these children could not be used to decide one way or the other.

A number of clinical protocols also recommend using total disfluencies as a metric for helping a clinician distinguish between a stuttering and nonstuttering child. Adams (1977) suggested that a child should be designated as stuttering when more than 10% of words spoken are disfluent. This principle has also been utilized by others (e.g., Gregory & Hill, 1993; Pindzola & White, 1986). Yairi (1997) also recommends use of this principle, despite the “nonstuttering” diagnosis applied to the child described above as Participant 15. Ambrose and Yairi (1999) extended the principle by emphasizing the proportions of SLDs in relation to other disfluencies when attempting to diagnose stuttering in children (see below).

Undoubtedly, stuttering children do, on average, exhibit more disfluencies than nonstuttering children. This was shown clearly within Johnson et al.’s (1959) comparison between speech samples from 89 stuttering children (68 boys, 21 girls; age = 29–96 months) and from a matched group of nonstuttering children. The mean total disfluencies for male stuttering speakers was 17.91 (number of nonfluencies per hundred words), and for the male controls it was 7.28. The females showed a similar group differences: 16.25 for the stuttering group and 7.90 for the control group (see Table 1). Table 1 also shows that essentially the same differences have occurred in all subsequent studies that have reported total disfluency counts for stuttering and nonstuttering children. The main difference between stuttering speakers and controls occurred in the categories that could be classified as SLDs, but not in their other disfluencies. The difference between the stuttering and nonstuttering children in the SLDs category range from 7.24 to 14.06, but in the other disfluencies category the differences only range from 2.5 (higher for stuttering children) to 4.2 (higher for nonstuttering children; see Table 1). In general, the findings show that stuttering speakers and controls do not differ when compared on other disfluency measures.

### ***Disfluencies That Characterize Stuttering and the Concept of an SLD***

In the above-mentioned study by Johnson et al. (1959), there were significant differences between the two groups

<sup>2</sup>The terms *disfluency*, *dysfluency*, and even *nonfluency* are frequently used as synonyms for stuttering (American Speech-Language-Hearing Association, 1999). This makes the terminology used in stuttering research unnecessarily confusing. The confusion is most apparent when it is not clear whether the user intended the word to be a descriptor for nonstuttered or for stuttered disfluencies produced by a stuttering speaker.

**TABLE 1. Studies that have reported data showing total disfluencies divided into stuttering-like disfluencies (SLDs) and other disfluencies.**

Study	Total		SLDs		Other disfluencies	
	Children who stutter	Controls	Children who stutter	Controls	Children who stutter	Controls
Johnson et al. (1959) <sup>a</sup> (words) Age: 2–8 years 68 M 21 F	17.91	7.28	11.51	1.88	6.40	5.40
	16.25	7.90	9.45	2.21	6.80	5.69
Yairi & Lewis (1984) <sup>b</sup> (syllables) Age: 2–3 years 10 P/10 C (5 M/5 F)	21.46	6.18	16.43	3.02	5.03	3.16
Hubbard & Yairi (1988) <sup>b</sup> (syllables) Age: 2–4 years 15 P/15 C (10 M/5 F)	22.45	5.89	16.88	2.82	5.57	3.07
Ambrose & Yairi (1999) <sup>b</sup> (syllables) Age: 2–5 years 90 P (59 M/31 F) 54 C (36 M/18 F)	15.78	5.65	10.37	1.33	5.41	4.32
Pellowski & Conture (2002) (words) Age: 3–4 years 36 P/36 C (25 M/11 F)	10.70	2.60	8.70	1.10	2.00	1.50
Zackheim & Conture (2003) <sup>b</sup> (words) Age: 3–5 years 6 P (6 M)/6 C (5 M/1 F)	12.70	5.40	12.00	0.50	0.70	4.90
Logan (2003) <sup>b</sup> (syllables) Age: 3–4 years 15 P/15 C (14 M/1 F)	9.64	3.30	7.74	1.50	1.90	1.80

*Note.* All data are mean number of disfluencies per 100 words or syllables. With the exception of Pellowski & Conture (2002), all data were recalculated by the present authors from information in the published article. Except where noted, SLDs comprised part-word repetitions, single-syllable word repetitions, and dysrhythmic phonation (tense pause). Other disfluencies consisted of interjections, revision/abandoned utterances, and multisyllable word and phrase repetitions. P = participants; C = controls; M = males; F = females.

<sup>a</sup>For this study, SLDs consisted of sound and syllable repetition, word repetition, broken words, and prolonged sounds, and other disfluencies consisted of interjections, phrase repetitions, revisions, and incomplete phrases. <sup>b</sup>For these studies, SLDs refer to repetitions/prolongations only. Data were averages of measures obtained during single and multiple utterance responses.

of children with respect to disfluencies categorized as sound and syllable repetitions, word repetitions, phrase repetitions, and prolonged sounds (plus total disfluencies: all  $p < .01$ ). Broken words also showed a significant difference ( $p < .05$ ). No significant differences were found for interjections, revisions, and incomplete phrases. Johnson and colleagues famously interpreted these findings as showing the overlap, and hence resemblances, between the disfluencies produced by stuttering and nonstuttering children, but this interpretation would have been untenable if the data had also included counts of perceived stuttering events. In fact, McDearmon's (1968) "re-examination of certain data" from the Johnson et al. (1959) study concluded that there was a "difference in kind" (p. 637) between the disfluencies of both groups of children. For some reason, though, this recognition of the fundamental differences between Johnson's disfluency typology and

stuttering behavior was ignored—and continues to be ignored. Arguably the direct result was that Johnson's disfluency categories gradually achieved an unjustifiable influence over stuttering behavior measurement.

Packman and Onslow (1998) noted that studies that had used Johnson's categories to describe stuttering often used language that appears to be inconsistent and imprecise. There is certainly plenty of empirical justification for their observation, given the poor reliability of judgments based on those categories (see below). The SLD metric, which was also based on Johnson's categories, may have also incorporated this imprecision. The choice of the SLD metric for the Illinois Studies was driven, according to Yairi (1997), by a mixture of tradition and studies comparing disfluencies in stuttering and nonstuttering children. Yairi also argued that because stuttering identification is often clouded by perceptual ambiguities

and theory, a disfluency-type measure (such as the SLD) is a less ambiguous and more atheoretical measure. However, because the SLD is constructed out of Johnson's categories, with their embedded imprecision and theoretic assumptions, that argument is unconvincing.

The issues surrounding the justification for using the SLD measure also incorporate another layer of complexity, because the disfluency types that constitute an SLD have varied greatly and have not always been documented clearly. Consider, for example, the constituents of an SLD as reported by Yairi in 1997: *part-word* and *monosyllabic word repetition*, *dysrhythmic phonation*, and *tense pause*. Some justification for using these terms to define SLDs comes from a study by Yairi and Lewis (1984), who used parent judgments to classify their children as having a stuttering problem or as being normally fluent. The children's speech samples were analyzed, and the disfluencies characterizing those with a stuttering problem were extracted. It was reported that only two disfluency types distinguished between their child groups: part-word repetitions and dysrhythmic phonation. The frequency of *single-syllable word repetitions* did not differentiate between the two groups (but the number of iterations per instance of repetition did). Nevertheless, monosyllabic repetitions were included as SLDs, a decision that continues to be an important consideration with respect to the use of SLDs. Yairi (1996) went to the hub of the issue by highlighting the lack of consistency among researchers in classifying a monosyllabic word repetition as an occasion of stuttering. Wingate (2001) remarked that monosyllabic word repetitions are usually considered to be common in normally fluent speech. This seems to be at odds with Ambrose and Yairi's (1999) and Yairi, Watkins, Ambrose, and Paden's (2001) conclusions that monosyllabic repetitions are distinctive in young stuttering speakers—presumably justifying their inclusion as an SLD.

The dysrhythmic phonation and tense pause labels for SLDs apparently emerged from Williams, Silverman, and Kools's (1968) modification of Johnson's (1961) system. A dysrhythmic phonation was defined as occurring only within words; similar behaviors that occurred between words (or between part-words or nonwords) were identified as tense pause (tension). Dysrhythmic phonation ultimately emerged as a label for a variety of disfluency subtypes. According to Yairi and Ambrose (1999, p. 1100), dysrhythmic phonation "includes sound prolongations, silent blocks and also broken words and other within-word interruptions (but not repetitions) that disturb the continuity of words." Ultimately, tense pause was omitted from the SLD metric because of the "difficulty in reliable identification" (Ambrose & Yairi, 1999, pp. 898–899), just as Johnson had observed 40 years earlier.

The changes that have been made over time to the composition of SLD units also make it difficult for researchers to know if the disfluency types that they are using as "SLDs" refer to the same unit. This problem, in turn, makes it difficult to understand the measurement methodology that is actually being employed within some recent studies. Consider, for example, recent studies by

Logan (2003) and Pellowski and Conture (2002); both use different sets of disfluencies within their "SLD" and "other disfluency" measures, yet both reference Ambrose and Yairi (1999) as the basis for their measures. Trautman et al. (2001) used yet another form of SLD, "stuttering-like disfluencies," but still claimed that they were derived from Yairi and Ambrose (1992). In fact, as Table 2 shows, we were able to identify 13 different disfluency-type systems that could be regarded as derivatives of that study.

Finally, it is noteworthy that in the latest Illinois Studies the SLD unit is linked to a word; it does not refer to events that occur between words (Ambrose & Yairi, 1999; Yairi & Ambrose, 1999). This is also puzzling because of Yairi's (1996) published acknowledgment that there is ample evidence confirming that stuttering events occur *between* as well as *within* word production (Cordes & Ingham, 1995b).<sup>3</sup>

### ***How Have SLDs Been Used to Classify Children Who Stutter?***

Yairi and colleagues have generally claimed that a child must display at least three SLDs per 100 syllables in a speech sample in order to qualify as a child who stutters (Yairi, 1997; Yairi & Ambrose, 1999). Similar criteria for classifying a child as stuttering have been employed outside of the Illinois Studies. For instance, Pellowski and Conture (2002) and Zackheim and Conture (2003) classified a child as a child who stutters if three or more SLDs per 100 words (not syllables) occur during a speech sample. Logan (2003) utilized a criterion of three disfluencies per 100 syllables, but in this case the investigator had also prejudged those disfluencies to be stuttering. It is difficult to understand why a disfluency count criterion would be used to identify a stuttering speaker (see Logan, 2003, p. 180) if the disfluency counts have already been judged to be stuttering events. On the other hand, some criterion seems logical—for if very few stuttering events characterize an individual's speech, then at some low frequency level it is likely that they will not constitute a serious problem behavior.

It is seemingly indisputable that not all SLDs are stuttering events. Single part-word repetitions, single-syllable word repetitions, or short prolongations typically occur within the speech of nonstuttering children, and listeners have no difficulty in classifying their speech as normally fluent. Yairi (1997) also recognized that an SLD is not a synonym for stuttering, remarking that "if disfluencies that are called 'stuttering' are counted in the speech of nonstuttering children" the result would "contradict common sense" (p. 51). Little wonder, therefore, that researchers have had difficulty in deciding just what an SLD unit is supposed to be measuring.

In their 1999 article, Yairi and Ambrose argued that the three-SLD threshold need not be surpassed in order for a speech sample to be judged as displaying stuttering;

<sup>3</sup>This is also a reason why recent stuttering measurement methods, such as the Stuttering Measurement System (R. J. Ingham et al., 1999), use perceptual judgments of stuttering across and beyond the syllable boundary and derive percentage syllables stuttered data in relation to the remaining nonstuttered syllables spoken.

**TABLE 2. Types of disfluency categories used in stuttering research studies.**

Type of study	Study										
	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	5	6	7	8	9	10	11
<b>Stutterings/SLD</b>											
Word repetition	•	•								•	
Sound and syllable repetition	•									•	
Sound prolongation	•							•			•
Broken words	•							•			
<i>Part-word repetition</i>		•	•	•	•	•	•	•	•		•
<i>Tense pause (tension)</i>		•	•	•	•			•	•		•
<i>Disrhythmic phonation</i>		•	•	•	•	•	•		•		
<i>Single-syllable word repetition</i>			•	•	•	•	•		•		•
<i>Whole-word repetition &gt; 3 iterations</i>								•			
<i>Phrase repetition &gt; 3 iterations</i>								•			
<i>Phrase repetition</i>										•	
<i>Prolongation-hard attack</i>											•
<i>Audible/inaudible sound prolongation</i>										•	
<i>Blocks</i>											•
<b>Other disfluencies</b>											
Phrase repetition	•		•	•	•	•			•		•
Interjections		•	•	•	•	•	•		•	•	
Incomplete phrases	•										
Revision	•	•								•	•
<i>Multi/polysyllable word repetition</i>			•	•	•	•			•		•
<i>Revision/incomplete phrase</i>			•	•	•	•			•		
<i>Interjection of sounds and syllables</i>		•									
<i>Phrase/multisyllable word repetition</i>								•			
<i>Linguistic nonfluency</i>								•			
<i>Maze</i>								•			

Note. The typology of Johnson et al. (1959) is shown nonitalicized. Later additions to Johnson et al.'s categories (that have been employed with all or part of Johnson et al.'s categories) are shown in italics. The disfluency types have been grouped into those that have been principally used to describe stuttering (stutterings/SLD) and those that have not been normally used to describe stuttering (other disfluencies). 1 = Johnson et al. (1959); 2 = Williams et al. (1968); 3 = Yairi & Lewis (1984); 4 = Hubbard & Yairi (1988); 5 = Yairi & Ambrose (1992); 6 = Yairi et al. (1993); 7 = Yairi & Ambrose (1999); 8 = Trautman et al. (2001); 9 = Pellowski & Conture (2002); 10 = Logan (2003); 11 = Anderson & Conture (2004).

<sup>a</sup>No specified distinction was made in Johnson et al. (1959), Williams et al. (1968), Yairi & Lewis (1984), or Hubbard & Yairi (1988) between SLDs/stutterings and other types of disfluencies.

nevertheless, they also accept that "a few young, normally fluent children, especially 2- to 3-year-olds, may exhibit up to 5 SLD" (1999, p. 1103). In other words, a child may exhibit as many as five SLDs per 100 syllables and yet *not* be classified as displaying stuttering. More importantly, though, Yairi and Ambrose concede that a "child cannot be classified as a 'stutterer' without having a minimal amount of *observed stuttering* [italics added]" (Yairi & Ambrose, 1999, p. 1103). In other words, as is true for Logan (2003), the decision as to whether a child should be classified as "stuttering" relies not on SLD counts but on "observed stuttering." The questionable value of the SLD concept was also highlighted in a recent commentary by Wingate (2001). In response, Yairi et al. (2001, p. 587) argued as follows: "We agree with Wingate that stuttering-like disfluency is not 'stuttering.' This is why we use the term the way we use it. We count disfluencies." One might reasonably ask, therefore, why SLDs would be favored over "observed stutterings" as the metric of choice in the Illinois Studies.

One method that Ambrose and Yairi (1999) have advanced for increasing the diagnostic strength of the SLD measure is to derive a *weighted SLD* score. But it is also a metric that requires considerable data analysis. Initially, the clinician must count and designate three types of

disfluencies as SLDs: part-word repetition, single-syllable word repetition, and dysrhythmic phonation (prolongations, blocks, and broken words), as well as repetition units (see p. 899). But *every* disfluency type must be tabulated as well because "the proportion of SLDs in children who stutter," it is asserted, "is 65% or higher of the total disfluency [count in their speech]" (p. 906). However, the ultimate differentiation between a stuttering and nonstuttering child requires that these data then be converted into a weighted score. This score is "determined by adding the part-word and single-syllable word repetition frequency per 100 syllables and multiplying by the mean number of repetition units. To this is added 2 times the frequency of dysrhythmic phonation: [(pw + ss) × ru] + (2 × dp)" (p. 899).

Ambrose and Yairi (1999, p. 906) claim that a weighted SLD score will "provide more powerful information about the presence and severity of stuttering." They also claim that a "weighted score of 4 or above should be suspected of stuttering regardless of the number of SLD" (p. 906). But these claims are not well supported. According to Ambrose and Yairi (2001, p. 596), the weighted score "is a mathematical method of showing what it is that lets us know one person is stuttering and another not." The problem

is that even after making this claim, Ambrose and Yairi (1999, p. 906) partially undercut its value because, as they state, the resulting score "is not to be taken as an absolute line of demarcation." This is probably not too surprising because, as previously noted, the weighted score is derived from data that have questionable validity (see also Onslow & Packman, 2001). Its value should also be reflected in usage. However, at the time of writing, not even the Illinois Studies have used this metric. In fact, only one study, by Pellowski and Conture (2002), reports using the weighted score. In that study the parents' judgments of children as potentially having a stuttering problem was used to verify whether the weighted score could distinguish these children (3–4 years old;  $n = 36$ ) from age-matched controls. The score did separate the groups in 97% of cases, but no evidence was provided to show that simply relying on the parents' judgments would not have been just as useful for distinguishing between these children.

### ***Do Different Disfluency Types Predict Persistence and Recovery From Chronic Stuttering?***

Another important reason that has been advanced for obtaining disfluency-type measures in stuttering speakers is that they are useful predictors of the development of stuttering. For instance, many authorities on stuttering have argued about the relative importance of prolongations (dysrhythmic speech) and repetitions in early stuttering. Van Riper (1971) believed, on the basis of his clinical experience, that if early stuttering was dominated by repetitions then the chances for recovery were greater than if it was dominated by blocks and prolongations. Curlee (1980) argued that children who continue to stutter will more frequently display disfluencies characterized by struggle, or tense and effortful articulatory attacks. Riley and Riley (1982) make similar assumptions. They contend that children are more likely to develop persistent stuttering if their speech is characterized by tense part-word repetitions and *inaudible and audible sound blocks*. Dickson (1971), using parent reports, found that if children's disfluencies were characterized by repetitions then they were more likely to recover from stuttering than if they were characterized by blocks.

Conture (1990) also argues that the risk of chronic stuttering will increase if a child's proportion of sound prolongations exceeds 25% of total disfluencies, a claim that appears to have emerged from a study by Schwartz and Conture (1988) on 43 young stuttering children (mean age = 5.11 years). The main finding of this study related to chronological age that was found *not* to be correlated with the number and variety of associated behaviors, number of sound prolongations, or stuttering frequency. It is not clear, therefore, how these findings could be used to justify claims made for the importance of the proportion of prolongations in a child who stutters.

In a later study, Pellowski and Conture (2002) addressed the possible wider importance of dysrhythmic phonations. They found a nonsignificant correlation ( $r = .38, p = .09$ ) between time since the onset of stuttering and the percentage of dysrhythmic phonations for 4-year-old

children (even lower correlations occurred among their 3-year-old children). The small number of participants ( $n = 16$ ) in this study, the nonsignificance of the reported correlations, and individual differences in the percentage of dysrhythmic phonation across a very wide range (0%–80%) make it difficult to interpret the usefulness of these data. The findings suggest that the frequency of dysrhythmic phonation in stuttering children will not necessarily increase as they become older, but it is not clear what value this holds for clinicians or researchers.

Repetitions do, however, seem to occur more frequently than prolongations among the disfluencies of young children who stutter (Ambrose & Yairi, 1999; Johnson et al., 1959; Yairi & Lewis, 1984). From one perspective, therefore, repetitions might seem to be more clinically significant than prolongations. Yaruss, LaSalle, and Conture (1998) surveyed the diagnostic records of 100 children (mean age = 54.7 months,  $SD = 12.2$  months) referred because of parent concern about possible stuttering. In 47% of these children, the predominant disfluency type was repetitions; prolongations characterized the speech of only 26% of these children. In this study, no significant correlations were found between participants' chronological age, time since onset, and any other measure of speech fluency, except for disfluency type (but see Cordes, 2000a, for a discussion of the methodological limitations in this retrospective study). When children who were estimated to have been stuttering for over a year were compared with those stuttering for less than a year, the former showed significantly more sound prolongations, while the latter group showed proportionately more repetitions. Yaruss et al. did note, however, that this was not true for all of their children and that the length of time a child had been stuttering was unrelated to a disfluency type. This inconsistency is highlighted by Zebrowski (1991), who found that audible sound prolongations, not repetitions, were the most frequently produced type of disfluency among a small group of stuttering children (3–5 years old;  $n = 10$ ).

There is also little evidence that particular types of disfluencies in stuttering children will help predict their recovery. Yairi, Ambrose, Paden, and Throneburg (1996) and Yairi and Ambrose (1999) found that stuttering severity at onset did not predict untreated recovery. However, these studies did not report separate disfluency-type data with respect to recovery. On the other hand, Yairi (1997) observed that a substantial number of the recovered children in the Illinois Studies did seem to exhibit prolongations and blocks during the first months of stuttering and that prolongations "may become a more significant factor after stuttering has persisted for a longer period of time" (Yairi, 1997, p. 70). But the evidence to date does not support claims for this suggested benefit of disfluency typing.

In general, it seems clear that the proportions of prolongations or repetitions in the speech of young children do not predict the ultimate severity of the disorder or the likelihood of chronicity. There is, therefore, no reason why clinicians should be advised that the proportion of disfluency types in a young child's stuttering has prognostic value.

## *The Reliability of SLD and Disfluency-Type Measures*

The reliability of a measure is "related to the general trustworthiness of obtained data" (Cordes, 1994, p. 265). It is now well established that this trustworthiness is influenced by many variables, including judgment method and judge training (see Cordes & Ingham, 1994). In stuttering research, reliability is usually determined by comparisons between independent judges' total counts of stuttering events or via event-by-event comparisons between their judgments. Low interjudge agreement for total counts of stuttering events and especially event-by-event analysis of stuttering judgments is commonly found in stuttering research. However, there is a paradox: In studies on stuttering treatment, the reported reliability data are generally satisfactory (see Cordes & Ingham, 1994). One possible explanation for this paradox is that treatment studies typically use fewer judges (usually only two) than formal investigations of stuttering judgments; treatment studies may also be using judgments that are peculiar to the treatment research center. Indeed, high intraresearch center reliability definitely does not mean high interresearch center reliability. Kully and Boberg (1988) first drew attention to the issue of poor stuttering judgment agreement across research centers. Subsequently, R. J. Ingham and Cordes (1992) showed that this may stem from institutional differences between clinical training facilities, and the problem was then dramatically highlighted in a study by Cordes and Ingham (1995a). They found that pairs of prominent researchers working in the same location displayed high agreement with each other, but their total counts were dramatically different from those reported by pairs of researchers working in other locations. It is this problem that has now made it abundantly clear that standardized stuttering measurement training systems are urgently needed for clinicians and researchers working with persons who stutter (see Cordes & Ingham, 1999).

SLD or disfluency-type measures differ from the methods studied by Kully and Boberg (1988) and Cordes and Ingham (1995a) in several ways, including some that rely on transcribed speech samples. Interestingly, there is surprisingly little research on the validity or reliability of disfluency data derived from transcribed speech samples. An accurate transcription of a suitable-size speech sample is a formidable task—and, incidentally, a task not conducive to routine clinical practice. Thus it is important to know the cost-benefit ratio for such a demanding undertaking. For a transcript to be valid, it must include all utterances including, for example, multiple repetitions. This creates a conundrum: a truly accurate transcript must include indicators that bias independent judgments from the same transcript. Conversely, if it omits such instances then, ipso facto, it is no longer a valid transcript. The issue is exemplified in a study by Hubbard (1998, p. 250), who provided judges with transcripts of speech samples from stuttering speakers that "consisted of only the words corresponding to the utterances." Unmentioned, however, was just how Hubbard verified that the selected words actually did correspond to counterpart utterances.

Some methods have been introduced to aid the transcription of disfluent events in a sample. One is use of the Systematic Analysis of Language Transcripts (SALT; Miller & Chapman, 1998) for transcribing speech samples from stuttering children (Logan, 2003; Trautman et al., 2001; Yairi & Ambrose, 1999). The SALT program is designed for language analyses, but it allows for coding of disfluencies. It is apparent, however, that some coding systems used with SALT will produce different disfluency counts across research groups. Consider the coding of a simple utterance such as "we are going." According to Yairi and Ambrose (personal communication, July 26, 2004), this utterance would be coded as: "[WW1] we are going" and would be classified as an SLD. For Logan (personal communication, July 26, 2004), it would be coded as "we we [Rep/pro] are going." Trautman et al. (2001) would consider the utterance to be a linguistic nonfluency because it is a whole-word repetition with less than three iterations. Suffice it to say, the need for a standardized system for coding disfluencies in stuttering research is obvious.

But not all disfluency-type measurement systems rely on transcriptions. Yaruss, Max, Newman, and Campbell (1998) report the use of a real-time technique for recording different disfluency types while watching a video sample. They report a comparison between the systematic disfluency analysis (SDA) transcription-based technique (Campbell & Hill, 1987) and this real-time technique that is described as follows:

Using a standard disfluency count sheet (e.g., Conture, 1990, p. 58; Yaruss, 1998), the clinician marks fluent words or syllables with a dash or a dot, while disfluent words or syllables are identified with an abbreviation indicating the type of disfluency produced (e.g., "SSR" for sound/syllable repetition). Next, the clinician calculates the overall frequency of disfluencies per 100 words or syllables produced, as well as the relative frequencies of various disfluency types. (Yaruss, Max, et al., 1998, p. 141)

In this report, only 200 syllables were used, although Conture (1990) has recommended that the samples should be at least 300 words. The procedure is recommended for use during treatment. The results of a comparison between the transcription and real-time recordings of the so-called "less typical" disfluencies showed that there were essentially trivial differences—pointing to high intergroup judgment agreement for total "less" and "more typical" disfluency counts.<sup>4</sup>

<sup>4</sup>Some issues raised by this study are particularly interesting. It is suggested, for instance, that clinicians using the SDA technique should not code disfluencies of less than 0.5-s duration because otherwise this would cause "overanalysis" of the speech performance. Nonetheless, it is concluded that the minor differences between the SDA and real-time technique probably occurred because the latter may have missed significant brief disfluencies. Suffice it to say that we are not aware of any data that would justify this reasoning. Indeed, it is simply not possible to know from this study's methodology whether the events identified during transcription or in real time had special claims to diagnostic accuracy.

In short, the findings highlight the utility of real-time judgments and that transcriptions are essentially redundant when measuring these types of disfluencies.

Studies that have used disfluency-type data have generally reported satisfactory inter- and intrajudge agreement on total disfluency counts. Yairi (1997), for instance, reviewed 11 studies that had used disfluency counts. They showed total count agreement scores that ranged from 0.62 to 0.92, and all except one study reported scores above 0.84. However, only two of these studies reported the reliability of judgments of particular disfluency types. The first is Johnson et al.'s (1959) study, but the data are not relevant because they were not derived from the study's participants.<sup>5</sup> The second is from a study by Hubbard and Yairi (1988); they reported only moderate reliability (0.59–0.67) for three of the four disfluency types that constitute an SLD.

Two studies have addressed the issue of the validity and reliability of using disfluency-type measures. The first, by Onslow, Gardner, Bryant, Stuckings, and Knight (1992), used five trained clinicians' judgments on 200 brief samples from stuttering and normal speaking children (2–4 years old). They found a high level of agreement among the clinicians in distinguishing between utterances that were judged stuttered and not stuttered. However, when the agreed stuttering samples were judged according to Johnson et al.'s (1959) disfluency types, there was no clear relationship between the disfluency types and stuttering events. Far more importantly, Onslow et al. broke from the traditional judgment instructions applied to disfluency-type measurement by allowing the judges the opportunity to assign one or more disfluency-type labels to the utterance. From 200 utterances that were classified as disfluent, the judges only agreed on the same category for 65 utterances. Of those 65 utterances, 36 were assigned the same single disfluency type, but another 29 were assigned multiple descriptors with at least one common descriptor assigned by all five judges. The most revealing finding was that the five clinicians assigned a common disfluency-type category to only 32.5% of the 200 utterances.

In a similar study by Cordes (2000b), 30 judges listened to brief intervals (5 s) of speech from different adult stuttering speakers. The judges were asked to assign one or more disfluency types from a list of 15 categories derived from three stuttering behavior definitions (Conture, 1990; Wingate, 1964; Yairi & Ambrose, 1992). Judgment agreement was determined by comparisons made between the judgments of 15 pairs of judges, all of whom listened several times to each interval before making a judgment. This task was repeated, and disagreeing judgments were resolved during discussion with a judge partner. A large number of findings emerged from this study, but in this context the most important general finding concerned the extent to which disfluency-type judgments could identify samples that authorities on stuttering had previously agreed were stuttered and nonstuttered

(Cordes & Ingham, 1995a). It was found that "there was no combination of disfluency types that identified more than 90.6% of the [agreed stuttering] intervals as containing stuttered disfluencies, and this definition [from Conture, 1990] resulted in up to 58% of the [agreed nonstuttered] intervals being labeled as containing stuttered disfluencies" (Cordes, 2000b, p. 959). This study, in conjunction with the Onslow et al. (1992) study, highlights the poor validity of stuttering-type labels for identifying speech that authoritative judges have agreed to be stuttered.

The Onslow et al. (1992) and Cordes (2000b) studies also show that the identification and quantification of disfluency types may, in fact, be far more problematic than judgments of stuttering events in stuttering speakers' speech samples. In this respect, therefore, they offer little support for claims that have been made for the advantages of disfluency-type measurement (see Yairi et al., 2001). Together, they also highlight a wider problem: the absence of any standardized method for training judges to identify the behaviors that correspond to the disfluency-type labels.

### *Have Disfluency-Type Measurements Made a Contribution to the Understanding and Treatment of Developmental Stuttering?*

Packman and Onslow (1998) have argued that an early treatment study by R. J. Ingham and Andrews (1971) constitutes a strong counterargument to the major thesis being developed here, or as evidence that the measurement of individual disfluency types can lead to knowledge that could not have been gained in any other way. The R. J. Ingham and Andrews study reported that there were differences between the types of "residual stuttering" that remained after a therapy program using prolonged speech and another using rhythmic speech. At the end of an intensive phase of therapy, the prolonged speech group was reported to show a much higher proportion of repetition-type stuttering, which was interpreted to mean that they had fewer severe occasions of stuttering and therefore a higher probability of maintained improvement at follow-up. Given the absence of reliability data for the repetition-type measures in the R. J. Ingham and Andrews study, it is difficult in retrospect to justify its conclusions. However, as later reported by R. J. Ingham and Andrews (1973), the decisive difference between these groups was that 9 months after treatment ceased, 15 of 23 (65%) of the prolonged speech group were stutter-free and 9 of 16 (56%) of the rhythmic treatment group were stutter-free. In other words, it was the superior overall speech performance of the prolonged speech group—not their disfluency topography—that led directly to prolonged speech being favored in later therapy programs investigated by R. J. Ingham and colleagues.

Another reasonable response to the question heading this section would be reference to the contributions from the Illinois Studies and their routine use of SLDs. Those studies have certainly had an important impact on the understanding of the likely course of developmental

<sup>5</sup>The data were obtained from independent judgments made on recordings from nonparticipant adult female stuttering speakers.

stuttering over the life span. And so, by implication, it could be argued that SLDs have aided that contribution. But was the SLD measure necessary to those findings, and, more importantly, could more knowledge have been gained by recording occurrences of stuttering in those studies? It will be recalled that the Illinois Studies employed two additional measures of speech, beside the SLD measure: ratings of stuttering severity by clinicians and by parents on a 0–7 rating scale. The latter two measures may in fact be the more important for the findings from this research. As Yairi et al. (2001, p. 586) acknowledge, “careful reading of the procedures [in the Illinois Studies] reveals that our main results—the percentage of recovery and persistency—would have remained essentially unaltered even if the SLD data were disregarded.”

Are there other important findings from the Illinois Studies that have benefited from the use of SLDs? The only obvious contender is the length of pause between a repeated part-word or a whole-word repetition. According to Throneburg and Yairi (1994), that interval was about 50% of the interval for normally fluent children and was strongly related to recovery. But the study’s most glaring omission was that it “did not employ direct correlations between segment duration and perceptual judgments of disfluent episodes as “normal” or “stuttering”” (Throneburg & Yairi, 1994, p. 1074). This was also true for a follow-up study (Throneburg & Yairi, 2001). In other words, perceptual judgments of “normal” or “stuttering” might have better estimated the probability of recovery. It remains to be shown, therefore, that the effort involved in measuring silent intervals in repetitions provides more information than could be gained by simply tracking the frequency of perceived stuttering events over the first 12 to 18 months after onset.

Another justification for the SLD measure is that it can be used to track the path of the disorder. For instance, the Illinois Studies reported that a stable level of SLDs 12 to 18 months after onset of stuttering may be a good predictor of chronicity, whereas a decline in SLDs by the end of the 1st year of stuttering strongly predicts recovery (Yairi, 1997; Yairi & Ambrose, 1992; Yairi et al., 1996). As yet, though, no published study has compared stuttering frequency measures and SLD measures for their predictive power. In fact, Ryan (2001) showed that recovery and persistent stuttering in young children could be predicted by monitoring the trend in stuttered words per minute scores. J. Ingham and Riley (1998) came to a similar conclusion by using a different metric (percentage syllables stuttered). *In other words, the prediction of persistence and recovery in young children can be made as easily and, arguably, more accurately by monitoring the trend of stuttering frequency during repeated assessments rather than by estimates of the number or proportion of disfluencies.* It is also important, as J. Ingham and Riley (1998) clearly showed, that speech performance measures obtained in a variety of speaking situations will indicate whether there is stability or valid signs of recovery.<sup>6</sup>

## Toward Some Solutions

One interesting recent attempt at improving the description accuracy of perceptually identified stuttering events has been to use a set of somewhat different descriptor labels. Packman and Onslow (1998) collapsed many of the well-established disfluency descriptors into three categories (“repeated movements,” “fixed postures,” and “superfluous behaviors”) to form what they termed the Lidcombe Behavioral Data Language (LBDL). But it is not yet clear that the methodology required to produce LBDL measurement warrants the required effort. An investigation by Teesson, Packman, and Onslow (2003) had both experienced and inexperienced judges repeatedly view 10-s speech samples; their findings revealed that their experienced judges did display higher interjudge agreement for particular LBDL-labeled behaviors occurring during the 10-s samples. Nonetheless, the findings do not show that exclusive LBDL labels were applied to individual stuttering events with high agreement. The latter issue also arose in a study by O’Brian, Packman, Onslow, and O’Brian (2004) that included use of LBDL measures of stuttering. In this instance, however, no LBDL judgment reliability data were reported. The “cash value” of LBDL labeling is whether it contributes more than the “stuttering” label to the diagnosis and treatment of stuttering. To date there is no evidence showing that it has any more value in this regard than, say, the venerable “clonic” and “tonic” labels for stuttering.

In the final analysis, perhaps Justice Potter Stewart’s famous failed attempt to define pornography—“I can’t define pornography, but I know it when I see it”—should apply to stuttering. Justice Stewart’s naive perspicacity might actually contain the seeds of a solution to the stuttering definition and measurement problem. If verbal descriptors of stuttering can never capture the features of the behavior that observers agree constitutes a stuttering event, then, as Cordes and Ingham (1994, 1999) have argued, perhaps we should try to capture what observers do agree on and then convert their “catch” into a measurement

---

<sup>6</sup>Arguably, there are some experimental investigations of stuttering behavior that may have contributed to the understanding of the disorder because of their use of disfluency-type measures. Martin and Siegel’s (1966) landmark study on the effects of response contingent shock on stuttering is a candidate study. They showed that when shock was delivered contingent on a “tongue protrusion” and after a while on an “S-prolongation,” then, for 1 participant, his “tongue protrusion” responses decreased but his “S-prolongation” response frequency was unaffected until shock was also delivered contingent on this response. It is not clear though that “tongue protrusion” corresponds to a disfluency response. The suggestion that different types of stuttering respond differently to contingencies is also not supported by Costello and Hurst (1981). They showed that when time out was delivered contingent on either Factor 1 or Factor 2 stuttering types (as defined by Bruten and Shoemaker, 1967), then both types of stuttering behavior decreased concurrently. Prins et al. (1980) may have also helped elucidate the effects on stuttering of the drug haloperidol when they noted that its “beneficial” effects did not extend to the disfluency types that typify stuttering but only occurred on nonstuttering-type disfluencies. However, the ineffectiveness of haloperidol in treating stuttering is well established in many other studies (see R. J. Ingham, 1984, chap. 11). In short, there is no compelling evidence among these studies that they derived new knowledge about stuttering as a result of quantifying the behavior in terms of disfluency types.

methodology. In other words, those instances of the behavior that appropriately qualified judges recognize and agree are stuttering events may offer a more fundamental basis for measuring stuttering and distinguishing between that behavior and normally disfluent speech. The idea of using an exemplar-based rather than verbal-based definition would not be a unique solution to a definition problem in our discipline. It appears to be the only solution open to differentially identifying some types of speech and voice disorders (see, e.g., Darley, Aronson, & Brown, 1969; Kreiman, Gerratt, Kempster, Erman, & Berke, 1993). Furthermore, it then provides an opportunity for more systematic investigations of the different parameters of the exemplars that contribute to their distinctiveness.

A program of research initiated in the 1990s by Cordes and Ingham has produced a corpus of 5-s audiovisual samples or exemplars from adult stuttering speakers that recognized authorities on stuttering consistently agree contain stuttering. These 5-s intervals are now embedded within a computer-based training program known as Stuttering Measurement Assessment and Training (R. J. Ingham, Cordes, Kilgo, & Moglia, 1998) that has been shown to improve not only the accuracy with which judges identify stuttering but also the interjudge agreement for stuttering event judgments (Cordes & Ingham, 1999). A similar approach is currently being developed for identifying stuttering in young children in this country (by A. Bothe) and in Iceland (by the first author). This reflects an important recent observation: that training to make stuttering judgments on adult or English speakers may not readily transfer to skillful judgments with children or foreign-language speakers. Nonetheless, this type of training has already provided the basis for training clinicians to use relatively simple computer programs to measure stuttering frequency in real time, along with measures of stutter-free speaking rate and, if necessary, speech naturalness (see R. J. Ingham, Bakker, Ingham, Kilgo, & Moglia, 1999).

The previously described approach need not be restricted to improving the measurement of agreed stuttering events; it could also be used to establish exemplar-based methods for measuring behaviors that relevant judges agree constitute SLDs or "more typical" and "less typical" disfluencies. Comparisons between and among such exemplar-based measures might then lead to research highlighting the extent of agreement/disagreement or redundancy among these measures. Cordes (2000b) has already provided a useful starting point for this type of research. Recently Van Borsel and Pereira (2005) made the interesting suggestion that exemplars of stuttering produced by speakers using different languages might increase the utility of exemplar-based training. At the very least, this strategy would help to ensure that different clinics or researchers who favor the investigation of these behaviors in the speech of persons who stutter are able to relate to a common set of referents. In turn, this would help to overcome the scientifically embarrassing problem of poor interclinic or interlaboratory stuttering measurement agreement (Cordes & Ingham, 1995a; R. J. Ingham & Cordes, 1992; Kully & Boberg, 1988).

## Conclusions

The heritage of disfluency-type measures, including "stuttering-like" or "stuttering-type" metrics, would strongly suggest that they were conceived in order to show the level of resemblance between stuttering and normally disfluent speech. Studies using these derived methods, as shown above, have continued to emphasize (usually unwittingly) this similarity. By continuing to use a disfluency-type measurement system, many studies have suggested that children who stutter are simply more disfluent, albeit in terms of the frequency with which they produce certain types of disfluencies. It is now almost indisputable that these children are not simply "more disfluent"; they are diagnosed with a speech disorder because they display *another* class of behaviors, one that only partially resembles normal disfluencies. It has been a category mistake of the highest order to conflate the class of behavior that constitutes stuttering with the class of behaviors that constitute normal disfluency. At the very least, there is no obvious justification for requiring clinicians to measure classes of speech behavior that are not considered to be problem behaviors.

## Acknowledgments

This article was prepared as part of the requirements for the first author's doctoral program at the University of Iceland and while visiting the University of California, Santa Barbara, during 2004. Special thanks are due to Dr. Anne Bothe for her very helpful comments and editing.

## References

- Adams, M. R. (1977). A clinical strategy for differentiating the normally nonfluent child and the incipient stammerer. *Journal of Fluency Disorders*, 2, 141-148.
- Ambrose, N., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research*, 42, 895-909.
- Ambrose, N., & Yairi, E. (2001). Perspectives on stuttering: Response to Onslow & Packman. *Journal of Speech, Language, and Hearing Research*, 44, 595-597.
- American Speech-Language-Hearing Association Special Interest Division 4: Fluency and Fluency Disorders. (1999, March). Terminology pertaining to fluency and fluency disorders: Guidelines. *ASHA*, 41(Suppl. 19), 29-36.
- Anderson, J. D., & Conture, E. G. (2004). Sentence structure priming in young children who do stutter and do not stutter. *Journal of Speech, Language, and Hearing Research*, 47, 552-571.
- Bloodstein, O. (1975). *A handbook on stuttering* (2nd ed.). Chicago: National Easter Seal Society.
- Bloodstein, O. (1995). *A handbook on stuttering* (5th ed.). San Diego, CA: Singular.
- Brutten, G. J., & Shoemaker, D. J. (1967). *The modification of stuttering*. Englewood Cliffs, NJ: Prentice-Hall.
- Campbell, J. H., & Hill, D. G. (1987, November). *Systematic disfluency analysis: Accountability for differential evaluation and treatment*. Miniseminar presented at the annual convention of the American Speech-Language-Hearing Association, New Orleans, LA.

- Campbell, J. H., & Hill, D. G.** (1994). Systematic disfluency analysis. In H. Gregory (Ed.), *Stuttering therapy: A workshop for specialists* (pp. 30–36). Evanston, IL: Northwestern University & The Speech Foundation of America.
- Carlo, E. J., & Watson, J. B.** (2003). Disfluencies of 3- and 5-year old Spanish-speaking children. *Journal of Fluency Disorders*, 28, 37–53.
- Conture, E. G.** (1990). *Stuttering* (2nd ed.) Englewood Cliffs, NJ: Prentice Hall.
- Conture, E. G.** (1997). Evaluating childhood stuttering. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2nd ed., pp. 239–257). Needham Heights, MA: Allyn & Bacon.
- Conture, E. G.** (2001). *Stuttering: Its nature, diagnosis, and treatment*. Needham Heights, MA: Allyn & Bacon.
- Cordes, A. K.** (1994). The reliability of observational data: I. Theories and methods for speech-language pathology. *Journal of Speech and Hearing Research*, 37, 264–278.
- Cordes, A. K.** (2000a). Comments on Yaruss, LaSalle, and Conture (1998). *American Journal of Speech-Language Pathology*, 9, 162–165.
- Cordes, A. K.** (2000b). Individual and consensus judgments of disfluency types in the speech of persons who stutter. *Journal of Speech, Language, and Hearing Research*, 43, 951–964.
- Cordes, A. K., & Ingham, R. J.** (1994). The reliability of observational data: II. Issues in the identification and measurements of stuttering events. *Journal of Speech and Hearing Research*, 37, 279–294.
- Cordes, A. K., & Ingham, R. J.** (1995a). Judgments of stuttered and nonstuttered intervals by recognized authorities in stuttering research. *Journal of Speech and Hearing Research*, 38, 33–41.
- Cordes, A. K., & Ingham, R. J.** (1995b). Stuttering includes both within-word and between-word disfluencies. *Journal of Speech and Hearing Research*, 38, 382–386.
- Cordes, A. K., & Ingham, R. J.** (1996). Disfluency types and stuttering measurement: A necessary connection? *Journal of Speech and Hearing Research*, 39, 404–405.
- Cordes, A. K., & Ingham, R. J.** (1999). Effects of time-interval judgment training on real-time measurement of stuttering. *Journal of Speech, Language, and Hearing Research*, 42, 862–879.
- Costello, J. M., & Hurst, M. R.** (1981). An analysis of the relationship among stuttering behaviors. *Journal of Speech and Hearing Research*, 24, 247–256.
- Curlee, R. F.** (1980). A case selection strategy for young disfluent children. In W. Perkins (Ed.), *Seminars in speech, language, and hearing* (pp. 277–287). New York: Thieme-Stratton.
- Darley, F. L., Aronson, A. E., & Brown, J. R.** (1969). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research*, 12, 462–496.
- Dickson, S.** (1971). Incipient stuttering and spontaneous remission of stuttered speech. *Journal of Communication Disorders*, 4, 99–110.
- Gordon, P. A., & Luper, H. L.** (1992). The early identification of beginning stuttering II: Problems. *American Journal of Speech-Language Pathology*, 1(3), 49–55.
- Gregory, H. H., & Hill, D.** (1993). Differential evaluation-differential therapy for stuttering children. In R. F. Curlee (Ed.), *Stuttering and related disorders of fluency* (pp. 23–44). New York: Thieme.
- Howell, P., Sackin, S., & Glenn, K.** (1997a). Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: I. Psychometric procedures appropriate for selection of training material for lexical dysfluency classifiers. *Journal of Speech, Language, and Hearing Research*, 40, 1073–1084.
- Howell, P., Sackin, S., & Glenn, K.** (1997b). Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II. ANN recognition of repetitions and prolongations with supplied word segment markers. *Journal of Speech, Language, and Hearing Research*, 40, 1085–1096.
- Hubbard, C.** (1998). Reliability of judgments of stuttering and disfluency in young children's speech. *Journal of Communication Disorders*, 31, 245–260.
- Hubbard, C. P., & Yairi, E.** (1988). Clustering of disfluencies in the speech of stuttering and nonstuttering preschool children. *Journal of Speech and Hearing Research*, 31, 228–233.
- Ingham, J., & Riley, G.** (1998). Guidelines for documentation of treatment efficacy for young children who stutter. *Journal of Speech, Language, and Hearing Research*, 41, 753–770.
- Ingham, R. J.** (1984). *Stuttering and behavior therapy: Current status and empirical foundations*. San Diego, CA: College-Hill Press.
- Ingham, R. J.** (1985). Toward a therapy assessment procedure for treating stuttering in children. In *Stuttering therapy: Prevention and intervention with children* (Publication No. 20, pp. 101–126). Memphis, TN: Speech Foundation of America.
- Ingham, R. J., & Andrews, G.** (1971). Stuttering: The quality of fluency after treatment. *Journal of Communication Disorders*, 4, 279–288.
- Ingham, R. J., & Andrews, G.** (1973). An analysis of a token economy in stuttering therapy. *Journal of Applied Behavior Analysis*, 6, 219–229.
- Ingham, R. J., Bakker, K., Ingham, J. C., Kilgo, M., & Moglia, R.** (1999). *Stuttering Measurement System (SMS)*. Santa Barbara: University of California, Santa Barbara.
- Ingham, R. J., & Bothe, A.** (2001). Recovery from early stuttering: Additional issues within the Onslow & Packman-Yairi & Ambrose (1999) exchange [Letter to the editor]. *Journal of Speech, Language, and Hearing Research*, 44, 862–867.
- Ingham, R. J., & Cordes, A.** (1992). Interclinic differences in stuttering event counts. *Journal of Fluency Disorders*, 17, 171–176.
- Ingham, R. J., Cordes, A. K., Kilgo, M., & Moglia, R.** (1998). *Stuttering Measurement Assessment and Training (SMAAT)* [Computer software and videodisk]. Santa Barbara: University of California, Santa Barbara.
- Johnson, W.** (1955). The time, the place, and the problem. In W. Johnson & R. R. Leutenegger (Eds.), *Stuttering in children and adults* (pp. 3–25). Minneapolis: University of Minnesota Press.
- Johnson, W.** (1956). Stuttering. In W. Johnson, S. J. Brown, J. J. Curtis, C. W. Edney, & J. Keaster (Eds.), *Speech handicapped school children* (Rev. ed., pp. 229–329). New York: Harper & Bros.
- Johnson, W.** (1961). Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers. *Journal of Speech and Hearing Disorders* (Monograph Suppl. 7), 1–20.
- Johnson, W., Boehmler, R. M., Dahlstrom, W. G., Darley, F. L., Goodstein, L. D., Kools, J. A., et al.** (1959). *The onset of stuttering*. Minneapolis: University of Minnesota Press.

- Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research, 36*, 21–40.
- Kully, D., & Boberg, E. (1988). An investigation of interclinic agreement in the identification of fluent and stuttered syllables. *Journal of Fluency Disorders, 13*, 309–318.
- Logan, K. (2003). Language and fluency characteristics of preschoolers' multiple-utterance conversational turns. *Journal of Speech, Language, and Hearing Research, 46*, 178–189.
- Martin, R. R., & Haroldson, S. K. (1981). Stuttering identification: Standard definition and moment of stuttering. *Journal of Speech and Hearing Research, 24*, 59–63.
- Martin, R. R., & Siegel, G. M. (1966). The effects of response contingent shock on stuttering. *Journal of Speech and Hearing Research, 9*, 340–352.
- McDearmon, J. R. (1968). Primary stuttering at the onset of stuttering: A re-examination of the data. *Journal of Speech and Hearing Research, 11*, 631–637.
- Miller, J. F., & Chapman, R. S. (1998). Systematic Analysis of Language Transcripts [Computer software manual]. Madison: University of Wisconsin—Madison, Waisman Research Center, Language Analysis Laboratory.
- O'Brian, S., Packman, A., Onslow, M., & O'Brian, N. (2004). Measurement of stuttering in adults: Comparison of stuttering-rate and severity-scaling methods. *Journal of Speech, Language, and Hearing Research, 47*, 1081–1087.
- Onslow, M. (1992). Identification of early stuttering: Issues and suggested strategies. *American Journal of Speech-Language Pathology, 1*(4), 21–27.
- Onslow, M., Gardner, K., Bryant, K. M., Stuckings, C. M., & Knight, T. (1992). Stuttered and normal speech events in early childhood: The validity of a behavioral data language. *Journal of Speech and Hearing Research, 35*, 79–87.
- Onslow, M., & Packman, A. (2001). Ambiguity and algorithms in diagnosing early stuttering: Comments on Ambrose and Yairi (1999). *Journal of Speech, Language, and Hearing Research, 44*, 593–594.
- Packman, A., & Onslow, M. (1998). The behavioral data language of stuttering. In A. K. Cordes & R. J. Ingham (Eds.), *Treatment efficacy for stuttering: A search for empirical bases* (pp. 27–50). San Diego, CA: Singular.
- Pellowski, M. W., & Conture, E. G. (2002). Characteristics of speech disfluency and stuttering behaviors in 3- and 4-year-old children. *Journal of Speech, Language, and Hearing Research, 45*, 20–34.
- Perkins, W. H. (1983). The problem of definition: Commentary on "stuttering." *Journal of Speech and Hearing Disorders, 48*, 246–249.
- Pindzola, R. H., & White, D. T. (1986). A protocol for differentiating the incipient stutterer. *Language, Speech, and Hearing Services in Schools, 17*, 2–15.
- Prins, D., Mandelkorn, T., & Cerf, F. A. (1980). Principal and differential effects of haloperidol and placebo treatments upon speech disfluencies in stutterers. *Journal of Speech and Hearing Research, 23*, 614–629.
- Riley, G. D., & Riley, J. (1982). Evaluating stuttering problems in children. *Journal of Childhood Communication Disorders, 6*, 15–25.
- Ryan, B. P. (2001). Prediction of spontaneous recovery in preschool children who stutter. In H. G. Bosshardt, J. S. Yaruss, & H. F. M. Peters (Eds.), *Proceedings of the Third World Congress on Fluency Disorders: Theory, Research, Treatment, and Self-Help* (pp. 206–210). Nijmegen, The Netherlands: Nijmegen University Press.
- Schwartz, H. D., & Conture, E. C. (1988). Subgrouping young stutterers: Preliminary behavioral observations. *Journal of Speech and Hearing Research, 31*, 62–71.
- Smith, A., & Kelly, E. (1997). Stuttering: A dynamic, multi-factorial model. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2nd ed., pp. 206–217). Needham Heights, MA: Allyn & Bacon.
- Teesson, K., Packman, A., & Onslow, M. (2003). The Lidcombe behavioral data language of stuttering. *Journal of Speech, Language, and Hearing Research, 46*, 1009–1015.
- Throneburg, R. N., & Yairi, E. (1994). Temporal dynamics of repetitions during the early stage of childhood stuttering—an acoustic study. *Journal of Speech and Hearing Research, 37*, 1067–1075.
- Throneburg, R. N., & Yairi, E. (2001). Durational, proportionate, and absolute frequency characteristics of disfluencies: A longitudinal study regarding persistence and recovery. *Journal of Speech, Language, and Hearing Research, 44*, 38–51.
- Trautman, L. S., Healey, E. C., Brown, T. A., Brown, P., & Jermano, S. (1999). A further analysis of narrative skills of children who stutter. *Journal of Communication Disorders, 32*, 297–315.
- Trautman, L. S., Healey, E. C., & Norris, J. A. (2001). The effects of contextualization on fluency in three groups of children. *Journal of Speech, Language, and Hearing Research, 44*, 564–576.
- Van Borsel, J., & Pereira, M. (2005). Assessment of stuttering in a familiar versus an unfamiliar language. *Journal of Fluency Disorders, 30*, 109–124.
- Van Riper, C. (1971). *The nature of stuttering*. London: Prentice Hall.
- Williams, D. E., Silverman, F. H., & Kools, J. A. (1968). Disfluency behavior of elementary school stutterers and nonstutterers: The adaptation effect. *Journal of Speech and Hearing Research, 11*, 622–630.
- Wingate, M. E. (1964). A standard definition of stuttering. *Journal of Speech and Hearing Disorders, 29*, 484–489.
- Wingate, M. E. (1987). Fluency and disfluency: Illusion and identification. *Journal of Fluency Disorders, 12*, 79–101.
- Wingate, M. E. (2001). SLD is not stuttering. *Journal of Speech, Language, and Hearing Research, 44*, 381–383.
- Yairi, E. (1982). Longitudinal studies of disfluencies in two-year-old children. *Journal of Speech and Hearing Research, 25*, 155–160.
- Yairi, E. (1996). Applications of disfluencies in measurements of stuttering. *Journal of Speech and Hearing Research, 39*, 402–403.
- Yairi, E. (1997). Disfluency characteristics of early childhood stuttering. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2nd ed., pp. 49–79). Needham Heights, MA: Allyn & Bacon.
- Yairi, E., & Ambrose, N. (1992). A longitudinal study of stuttering in children: A preliminary report. *Journal of Speech and Hearing Research, 35*, 755–760.
- Yairi, E., & Ambrose, N. (1999). Early childhood stuttering I: Persistency and recovery rates. *Journal of Speech, Language, and Hearing Research, 42*, 1097–1112.
- Yairi, E., Ambrose, N., & Niermann, R. (1993). The early months of stuttering: A developmental study. *Journal of Speech and Hearing Research, 36*, 521–528.
- Yairi, E., Ambrose, N., Paden, E., & Throneburg, R. (1996). Predictive factors of persistence and recovery: Pathways of childhood stuttering. *Journal of Communication Disorders, 29*, 53–77.

- Yairi, E., & Lewis, B.** (1984). Disfluency near the onset of stuttering. *Journal of Speech and Hearing Research*, 27, 154–159.
- Yairi, E., Watkins, R., Ambrose, N., & Paden, E.** (2001). What is stuttering? *Journal of Speech, Language, and Hearing Research*, 44, 585–597.
- Yaruss, J. S.** (1997). Clinical measurement of stuttering behaviors. *Contemporary Issues in Communication Science and Disorders*, 24, 33–44.
- Yaruss, J. S.** (1998). Real-time analysis of speech fluency: Procedures and reliability training. *American Journal of Speech-Language Pathology*, 7(2), 25–37.
- Yaruss, J. S., LaSalle, L. R., & Conture, E. G.** (1998). Evaluating stuttering in children: Diagnostic data. *American Journal of Speech-Language Pathology*, 7(4), 62–76.
- Yaruss, J. S., Max, M. S., Newman, R., & Campbell, J. H.** (1998). Comparing real-time and transcript-based techniques for measuring stuttering. *Journal of Fluency Disorders*, 23, 137–151.
- Zackheim, C. T., & Conture, E. G.** (2003). Childhood stuttering and speech disfluencies in relation to children's mean length of utterance: A preliminary study. *Journal of Fluency Disorders*, 28, 115–141.
- Zebrowski, P. M.** (1991). Duration of the speech disfluencies of beginning stutterers. *Journal of Speech and Hearing Research*, 34, 483–491.
- Zebrowski, P. M., & Kelly, E. M.** (2002). *Manual of stuttering intervention*. New York: Thomson Singular.

Received October 26, 2004

Accepted April 19, 2005

DOI: 10.1044/1058-0360(2005/026)

Contact author: Roger J. Ingham, Department of Speech and Hearing Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106. E-mail: rjingham@speech.ucsb.edu