Clinical Measurement of Stuttering Behaviors

J. Scott Yaruss
Northwestern University, Evanston, IL

Several recent studies have revealed that speech-language pathologists are generally less comfortable diagnosing and treating stuttering than other speech and language disorders (Cooper & Cooper, 1985, 1996; Mallard, Gardner, & Downey, 1988; Sommers & Caruso, 1995; St. Louis & Durrenberger, 1993; St. Louis & Lass, 1981). One reason for this may be the fact that there has been considerable debate among researchers regarding the appropriate techniques for measuring stuttering. Some recent topics of discussion have included: (a) the unit of measurement (instances of stuttering vs. intervals of time), (b) the specific behaviors to be measured (only stuttered disfluencies or a broad range of disfluency types), (c) the classification of disfluency types, and (d) the reliability and validity of various measurement techniques. These debates can greatly improve our understanding of the nature of stuttering; however, they can also leave students and clinicians confused and frustrated concerning which measurements are appropriate and how they should be made.

This confusion, combined with the recent elimination of the American Speech-Language-Hearing Association's (ASHA's) clinical training requirements for stuttering, can have serious consequences for speech-language pathologists working with people who stutter. Some student-clinicians may not receive adequate training concerning stuttering or sufficient opportunity to practice measuring stuttering while in graduate school. Others may be less consistent about making measures because they are uncertain about which measures should be made or how frequently they should be made. Still others may be reluctant to measure stuttering at all, arguing that "no one really knows how to measure stuttering anyway" or asking, "why measure stuttering today if it's only going to be different tomorrow?" (e.g., Milisen, 1971). Certainly, these are valid concerns; however, it is crucial for speech-language pathologists to measure their clients' stuttering behaviors to determine which clients need treatment and to document changes in their clients' behavior associated with treatment.

The purpose of this article is to attempt to reduce some of this confusion by discussing several issues related to the measurement of stuttering in children and adults. Specifically, this article will focus on issues related to the appropriate techniques for counting stuttering behaviors in the hopes that speech-language pathologists will become less apprehensive regarding their ability to accurately and appropriately measure the speech disfluency behaviors exhibited by their clients who stutter.¹

¹ The article will focus on the measurement of stuttering. For general discussions of the diagnostic process, readers should consult the extensive stuttering diagnostic literature (e.g., Adams, 1980; Conture, 1990b, 1997; Conture & Caruso, 1987; Conture & Yaruss, 1993; Costello & Ingham, 1984; Culatta & Goldberg, 1995; Curlee, 1993; Gordon & Luper, 1992a, 1992b; Gregory & Hill, 1993; Hill, 1995; Manning, 1996; Pindzola & White, 1986; Silverman, 1996a; Zebrowski, 1994).
BASIC QUESTIONS REGARDING STUTTERING MEASUREMENT

Why Measure Stuttering?

One of the first questions that may occur to a clinician thinking about the problems noted above is, “Why should I measure stuttering at all?” Though somewhat bold, this is a valid question. For a behavior as variable as stuttering, it is appropriate to question whether accurate and reliable measurements can be made. To begin, during a diagnostic evaluation, clinicians need to carefully measure a client’s speech fluency (and other behaviors) in order to make an appropriate judgment of whether the client is stuttering, whether the client will require treatment, and what the nature of that treatment should be. Later, during treatment, clinicians need to measure changes in their client’s speech to document their response to treatment and to help with planning future treatment. Clinicians need to be able to distinguish day-to-day or situation-to-situation fluctuations from changes associated with treatment. Although clinical intuition and experience will certainly play a role in these judgments, such decisions cannot and should not be made solely on the basis of “gut-level” feelings or rough guesses. Instead, clinicians should be able to consult their clinical records to determine whether there is a trend toward increased or decreased fluency across treatment sessions.

What Behaviors Should I Measure?

Once a clinician has decided in favor of consistently measuring stuttering, the next question may be, “What behaviors should I measure?” Unfortunately, this is one of the more difficult questions to answer, given the broad range of behaviors associated with stuttering (see Bloodstein, 1995, Chapter 1; Silverman, 1996b, Chapter 2). Stuttering affects many aspects of a client’s life other than speech, so there may be many cognitive and affective consequences of stuttering, such as low self-esteem, negative attitudes toward speaking, and feelings of shame or inadequacy. Accordingly, clinicians must be prepared to consider more than just observable speech behaviors when assessing individuals who stutter. A partial list of behaviors that can be measured is presented in Table 1.

The specific measures to be made will depend, in part, on whether the client’s speech is being assessed as part of a diagnostic evaluation or as part of an ongoing treatment program. During a diagnostic evaluation, the clinician will need to complete a broad and detailed analysis of the frequency and types of disfluency, as well as many other factors, such as how the client’s speech and language production affect his or her stuttering, the client’s attitudes toward speaking and reactions to stuttering, the effect of different speaking situations on stuttering, etc. During a treatment session, on the other hand, the clinician will need to focus more on the measures related to the specific treatment approach, such as the use of modification techniques, the level of speech naturalness, or the amount of physical tension associated with stuttering. Furthermore,

<table>
<thead>
<tr>
<th>Table 1. Selected aspects of client’s behavior that can be measured.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics of speech disfluencies</td>
</tr>
<tr>
<td>Frequency of disfluencies</td>
</tr>
<tr>
<td>Types of disfluencies (and combinations of disfluency types)</td>
</tr>
<tr>
<td>Duration of various types of disfluencies</td>
</tr>
<tr>
<td>Number or rhythm of iterations during repetitions</td>
</tr>
<tr>
<td>Audible or visible tension during disfluencies</td>
</tr>
<tr>
<td>Nonspeech behaviors associated with disfluencies</td>
</tr>
<tr>
<td>“Severity” of individual instances of disfluency</td>
</tr>
<tr>
<td>Factors that may affect the occurrence of disfluencies</td>
</tr>
<tr>
<td>Linguistic contexts in which disfluencies occur</td>
</tr>
<tr>
<td>Situational contexts in which disfluencies occur</td>
</tr>
<tr>
<td>Overall speaking rate (i.e., number of words produced per minute of talking time)</td>
</tr>
<tr>
<td>Articulatory speaking rate (i.e., number of syllables produced per second, with pauses, hesitations, and disfluencies removed)</td>
</tr>
<tr>
<td>Response time latency (i.e., the pause between the end of the conversational partner’s utterance and the beginning of the client’s utterance)</td>
</tr>
<tr>
<td>Client factors</td>
</tr>
<tr>
<td>Reaction to disfluencies or different speaking situations</td>
</tr>
<tr>
<td>Attitudes and feelings about speaking, stuttering, and self</td>
</tr>
<tr>
<td>Avoidance of sounds, words, or speaking situations</td>
</tr>
<tr>
<td>Response to treatment</td>
</tr>
<tr>
<td>Use of modification techniques</td>
</tr>
<tr>
<td>Effect of modification techniques on speech disfluencies</td>
</tr>
<tr>
<td>“Naturalness” of fluent speech</td>
</tr>
</tbody>
</table>

Note: This table uses the generic term “disfluencies” to refer to both stuttered and normal types of disfluencies. Measures can be made relative to instances of stuttering if preferred.

some behaviors will be more relevant for some clients than for others. For example, young clients may exhibit relatively little physical tension during disfluencies, so detailed assessments of tension may provide little meaningful information. Accordingly, the specific behaviors to be measured will vary depending on who the client is, the client’s reaction to stuttering, and where the client is in the treatment process.

How Often Should I Measure?

Once a client is in treatment, another important question is, “How often should I measure?” This is another potentially confusing issue because stuttering has been shown to vary from one testing session to another (Guitarrez & Caruso, 1995; Hillis, 1993; Ingham, 1985; Sheehan, 1969). Also, the answer to this question is highly dependent on the specific measure under consideration and the specific treatment approach being used.

Certainly, one of the most common measures to be made on a session-by-session basis will be a count of the number of disruptions in the client’s speech (e.g., Hillis, 1993; Ingham, 1985). Indeed, for some treatment approaches, regular assessment of the frequency of disfluencies is imperative (e.g., Boberg & Kully, 1985; Neilson &
Andrews, 1993; Ryan, 1979). For other treatment approaches, however, it may be less important that the frequency of disfluencies be assessed every session.²

Of course, the frequency and type of disfluencies are not the only measures of interest. How often a clinician measures other aspects of a client’s stuttering behaviors will depend on factors such as the client’s stage in treatment and rate of change, and the specific behaviors to be measured. For example, when a client is just beginning to learn a new modification technique, it may be appropriate to assess the client’s use of that modification during every session. Later, when the client becomes more adept at using the modification, it may be appropriate to assess the modification on a less frequent basis. Or, if the clinician is helping a client improve his attitudes toward speaking, it may be appropriate to examine changes in speech attitudes on a monthly basis, whereas attitudes may be measured less frequently when they are not the primary focus of treatment. Thus, the clinician should develop a flexible schedule in which some measures (e.g., frequency and type of disfluencies or use of modification techniques) are made on a regular basis and others (e.g., speech attitudes or linguistic contexts affecting fluency) are made on a less regular, but still consistent basis based on the specific needs of the client.

In What Situations Should I Measure?

Variability is one of the hallmarks of stuttering (e.g., Bloodstein, 1995; Conture, 1990a; Starkweather, 1987). Accordingly, it is imperative that the clinician obtain a representative sample of the client’s speech behaviors—both during diagnostic testing and throughout treatment. However, as shown in Table 2, a variety of factors can affect stuttering, such as the speaking task (i.e., the activity in which the speaker is engaged), the conversational setting (i.e., the location in which the speaking is taking place), and the conversational partner (i.e., the person with whom the client is speaking). Thus, it would seem quite unlikely that a representative sample can be collected in a single speaking situation (Costello & Ingham, 1984; Manning 1996; Yaruss, submitted). Accordingly, another important question a clinician should ask is, “In what situations should I measure?”

For adults who stutter, some speaking situations that have been commonly assessed include: (a) a reading task; (b) a picture description task using pictures such as those found in the Thematic Apperception Test (Murray, 1943); and (c) a monologue, such as having the speaker describe his or her job (e.g., Gregory, 1972; Johnson, 1961; Sander, 1961; Silverman, 1990b). In addition, it can be helpful to assess the relative influence of different speaking situations on adults’ stuttering by obtaining a “situational anxiety hierarchy,” or a rank-ordering of which situations are more or less difficult (e.g., Bruten & Shoemaker, 1967; Darley & Spriestersbach, 1978; Shumak, 1955).

For children who stutter, there has been less emphasis in the diagnostic literature on assessing situational variability. Although some diagnostic protocols have stressed the importance of evaluating different speaking situations (e.g., Campbell & Hill, 1987; Costello & Ingham, 1984; Gregory & Hill, 1993), many do not (see review in Gordon & Luper, 1992a, b). Still, in one recent study (Yaruss, submitted), children who stutter were shown to exhibit highly individualized patterns of variability across five different speaking situations (play, play with conversational pressures imposed, parent–child interaction, story retell, and picture description). Thus, it would seem appropriate for clinicians to select a standard set of speaking situations, both in and out of the clinic and with a variety of speaking partners, and supplement those with situations that may be appropriate for a specific client (e.g., Costello & Ingham, 1984). Obtaining such information can help a clinician determine what factors affect a client’s speech fluency, identify areas that may need further assessment or treatment, and better evaluate whether changes in fluency are associated with successful treatment.

### Table 2. Selected speaking tasks and situations that may affect a client’s speech fluency.

<table>
<thead>
<tr>
<th>Conversation tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
</tr>
<tr>
<td>Describing a picture</td>
</tr>
<tr>
<td>Speaking in a monologue</td>
</tr>
<tr>
<td>Speaking in a conversational dialogue</td>
</tr>
<tr>
<td>Speaking in a conversational dialogue with increased conversational pressures</td>
</tr>
<tr>
<td>Retelling a story</td>
</tr>
<tr>
<td>Describing job (for adults)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conversational settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the clinic</td>
</tr>
<tr>
<td>At home</td>
</tr>
<tr>
<td>In the classroom (for children)</td>
</tr>
<tr>
<td>On the playground (for children)</td>
</tr>
<tr>
<td>In the workplace (for adults)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conversational partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician</td>
</tr>
<tr>
<td>Peers</td>
</tr>
<tr>
<td>Parents (for children)</td>
</tr>
<tr>
<td>Siblings (for children)</td>
</tr>
<tr>
<td>Spouses (for adults)</td>
</tr>
<tr>
<td>Coworkers (for adults)</td>
</tr>
</tbody>
</table>

### BASIC ISSUES IN STUTTERING MEASUREMENT

#### Counting Disfluencies or Stuttering

In addition to these more general questions concerning stuttering measurement, clinicians are likely to encounter a number of other potentially confusing questions, such as what is the appropriate unit of measurement for counting stuttering. At first glance, it may seem obvious that this

---

² It is important to note that although positive change during treatment is often indicated by a decrease in the frequency of disfluencies, progress can also be indicated by an increase in the frequency of disfluencies if word substitution or avoidance of stuttering is decreasing.
would be instances of stuttering. As with other aspects of this particular speech disorder, however, it is not that simple. Indeed, there has been considerable debate among clinical researchers about whether it is appropriate to count instances of stuttering or instances of disfluency (e.g., Conture, 1990b; Cordes & Ingham, 1996a; Costello & Ingham, 1984; Ham, 1989; Ingham, 1985; Kully & Boberg, 1988).3

In general, those who prefer to count disfluencies argue that it is important to consider a broad range of disfluent behaviors when making diagnostic decisions (e.g., Conture, 1990b; Campbell & Hill, 1987; Gregory, 1986; Gregory & Hill, 1993; Webster, 1979), because the specific nature of the disfluencies can provide important diagnostic information (Adams, 1980; Conture, 1990b; Curlee, 1993; Riley, 1981, 1994; Schwartz & Conture, 1988; Yaruss, LaSalle, & Conture, submitted; Zebrowski & Conture, 1989). For example, a child who is primarily producing sound prolongations is often presumed to be at greater risk for continuing to stutter than a child who is primarily producing repetitions (e.g., Conture, 1990b).

By examining a variety of disfluency types, a clinician can also assess the total impact of disfluencies on a client’s overall communication effectiveness. In addition, some have suggested that the overall frequency of all disfluency types, not just those presumed to be stuttering, may help with the identification of children who stutter (e.g., Adams, 1980). Finally, because many different definitions of stuttering have been proposed (see Wingate, 1964, for a discussion of various criteria for defining stuttering), some have argued that it may be difficult, if not impossible, to agree on precise criteria for counting stuttering.

Those who favor counting stuttering, on the other hand, argue that stuttering is easily identifiable (e.g., Hamre, 1992; Ryan, 1974) and more relevant to the specific speech disorder under consideration (Costello & Ingham, 1984; Ingham, 1985). In other words, they might say that it makes little sense to measure interjections if these are a normal aspect of speech production and not characteristic of the individual’s stuttering disorder.

**Types of disfluencies.** If one chooses to count instances of disfluency, it is important to recognize that there are many different types of disfluencies that a speaker can produce. One of the first classification schemes was proposed by Johnson (1961; Johnson, Darley, & Spreijstersbach, 1963), who categorized disfluencies as either interjections, part-word repetitions, word repetitions, phrase repetitions, revisions, incomplete phrases, broken words, or prolonged sounds. This classification scheme has been modified somewhat by other writers (e.g., Sander, 1961; Williams, Silverman, & Kools, 1968; Wingate, 1976), for example, to more specifically differentiate types of part-word repetitions. In general, however, this method of classifying disfluencies has remained in use by many clinicians and researchers for years (Cordes & Ingham, 1994a; Lewis, 1991; Silverman, 1996a).

Although Johnson discussed differences in the types of disfluencies produced by speakers who do and do not stutter (Johnson & Associates, 1959), he primarily used this classification scheme to emphasize the considerable overlap in the types of disfluent behaviors produced by individuals who do and do not stutter (see also Bjorken, 1980; Yairi, 1972, 1981, 1982; Yairi & Lewis, 1984). Since that time, others have emphasized that findings that individuals who stutter are more likely to produce repetitions of sounds, syllables, or monosyllabic whole words; audible or inaudible sound prolongations; and dysrhythmic phonation, whereas individuals who do not stutter are more likely to produce interjections, revisions, and phrase repetitions (e.g., DeJoy & Gregory, 1985; Johnson & Associates, 1959; Yairi & Lewis, 1984). These findings, combined with the results of perceptual studies indicating that certain disfluency types are more likely to be perceived as stuttered by listeners (Boehm, 1958; Schiavetti, 1975; Williams & Kent, 1958; Zebrowski & Conture, 1989), have resulted in behavioral classification schemes designed to identify those disfluencies that are most likely to be instances of stuttering (see Table 3). For example, based on the observation that many of the disfluencies exhibited by individuals who stutter involve an interruption within a word unit (e.g., repetitions of sounds or syllables, but not phrases), one widely used behavioral categorization scheme separates disfluencies into “within-word” (stuttered) and “between-word” (nonstuttered) varieties (Conture, 1990a, 1990b).4

This kind of behavioral distinction can provide a useful, empirically derived guideline for describing and categorizing the types of disfluencies typically produced by individuals who stutter. For example, a recent analysis of the disfluencies exhibited by 100 young children who stutter (Yaruss, LaSalle, & Conture, submitted) revealed that over 87% of the disfluencies exhibited at the time of their initial diagnostic evaluations were within-word disfluencies. Still, not all within-word disfluencies are stuttered, and not all between-word disfluencies are nonstuttered. Therefore, clinicians must be cautious about relying too heavily on such behavioral guidelines for identifying instances of stuttering in their clients’ speech.

---

3 This distinction is related to fundamental questions regarding the nature and definition of stuttering, such as whether stuttering should be defined on a behavioral or perceptual basis (Conture, 1990a; Martin & Haroldson, 1981; Wingate, 1964; Young, 1984) and whether “normal” and “stuttered” disfluencies lie on a continuum or are categorically different events (Conture, 1990b; MacDonald & Martin, 1973; Young, 1984). Because this article is limited to discussing the clinical measurement of stuttering, these issues will not be further explored. Interested readers will gain a more thorough understanding of issues related to stuttering measurement by consulting the references cited above.

4 See Conture & Caruso (1987) and Conture (1990a, 1990b) for discussions of the role of probability in judgments in the identification of stuttering.

5 The strong form of this distinction—that all within-word disfluencies are stuttered and all between-word disfluencies are nonstuttered—has recently received some appropriate criticism (Cordes & Ingham, 1995b, 1996b; Yairi, 1996). It would appear, however, that most clinicians and researchers who use this within-word/between-word distinction do not ascribe to the strong form (see Conture, 1990a, who suggests that a consensus definition of stuttering may be helpful for both clinical practice and research, but who also discusses some of the liabilities of making such a strong behavioral differentiation between stuttered and normal disfluencies).
Table 3. Several ways of categorizing disfluencies.

<table>
<thead>
<tr>
<th>Within-word vs. between-word disfluencies</th>
<th>Between-word disfluencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e.g., Conture, 1990a, 1990b)</td>
<td></td>
</tr>
<tr>
<td><strong>Within-word disfluencies</strong></td>
<td><strong>Between-word disfluencies</strong></td>
</tr>
<tr>
<td>Monosyllabic whole-word repetition</td>
<td>Phrase repetition</td>
</tr>
<tr>
<td>Sound/syllable repetition</td>
<td>Polyisyllabic whole-word repetition</td>
</tr>
<tr>
<td>Audible prolongation</td>
<td>Interjection</td>
</tr>
<tr>
<td>Inaudible prolongation</td>
<td>Revision</td>
</tr>
<tr>
<td><strong>Stuttering-like disfluencies</strong></td>
<td></td>
</tr>
<tr>
<td>(e.g., Yairi, 1996; Yairi &amp; Ambrose, 1992; Yairi, Ambrose, &amp; Niermann, 1993; Yairi, Ambrose, Paden, &amp; Throneburg, 1996)</td>
<td></td>
</tr>
<tr>
<td>Stuttering-like disfluencies (SLD)</td>
<td></td>
</tr>
<tr>
<td>Part-word repetition</td>
<td>Interjection</td>
</tr>
<tr>
<td>Monosyllabic word repetition</td>
<td>Phrase repetition</td>
</tr>
<tr>
<td>Dysrhythmic phonation</td>
<td>Revision</td>
</tr>
<tr>
<td></td>
<td>Incomplete phrase</td>
</tr>
<tr>
<td><strong>Stutter-type disfluencies</strong></td>
<td></td>
</tr>
<tr>
<td>(e.g., Meyers, 1986)</td>
<td></td>
</tr>
<tr>
<td>Stutter-type disfluencies</td>
<td></td>
</tr>
<tr>
<td>Part-word repetition</td>
<td>Normal-type disfluencies</td>
</tr>
<tr>
<td>Prolongation</td>
<td>Whole-word repetition</td>
</tr>
<tr>
<td>Broken word</td>
<td>Phrase repetition</td>
</tr>
<tr>
<td>Tense pause</td>
<td>Revision</td>
</tr>
<tr>
<td></td>
<td>Incomplete phrase</td>
</tr>
<tr>
<td></td>
<td>Interjection</td>
</tr>
<tr>
<td><strong>Less-typical vs. more-typical disfluencies</strong></td>
<td></td>
</tr>
<tr>
<td>(e.g., Campbell &amp; Hill, 1987; Gregory, 1986, 1993)</td>
<td></td>
</tr>
<tr>
<td>Less-typical disfluencies</td>
<td>More-typical disfluencies</td>
</tr>
<tr>
<td>Monosyllabic word repetition</td>
<td>Hesitation</td>
</tr>
<tr>
<td>(3 or more repetitions)</td>
<td>Interjection</td>
</tr>
<tr>
<td>Part-word syllable repetition</td>
<td>Revision</td>
</tr>
<tr>
<td>(3 or more repetitions)</td>
<td>Phrase repetition</td>
</tr>
<tr>
<td>Sound repetition</td>
<td>Monosyllabic word repetition</td>
</tr>
<tr>
<td>Prolongation</td>
<td>(two or fewer repetitions; no tension)</td>
</tr>
<tr>
<td>Block</td>
<td>Part-word syllable repetition</td>
</tr>
<tr>
<td></td>
<td>(two or fewer repetitions; no tension)</td>
</tr>
</tbody>
</table>

**Note:** In each case, the first column represents those disfluencies judged to be more characteristic of individuals who do stutter and the second column represents those disfluencies judged to be more characteristic of individuals who do not stutter.

In an attempt to avoid the pitfalls of strict behavioral definitions, some researchers have used terms such as “stuttering-like” (Yairi, 1996; Yairi & Ambrose, 1992; Yairi, Ambrose, & Niermann, 1993; Yairi, Ambrose, Paden, & Throneburg, 1996), “less typical” (Campbell & Hill, 1987; Gregory, 1986; Gregory & Hill, 1993), or “stutter-type” (Meyers, 1986) to describe those disfluencies that are more characteristic of individuals who stutter. As can be seen in Table 3, these classification schemes are generally quite consistent in the behaviors labeled as similar to stuttering, though there are some interesting differences (particularly related to the treatment of monosyllabic whole-word repetitions; see Yairi, 1996).

**Disfluency clusters.** Categorizing disfluencies according to a set of unitary types such as those described above is not without its problems. Many times, more than one type of disfluency may be present in a given speech disruption, so it can be difficult to assign a single category (Boehmler, 1958; Johnson, 1961; Onslow, 1996; Onslow, Gardner, Bryant, Stuckings, & Knight, 1992; Young, 1961). For example, there are elements of both repetition and prolongation in a child’s production of a disfluency such as “m- mmmommy.” Such “multicomponent” (Campbell & Hill, 1987; Gregory, 1986) or “clustered” (Hubbard & Yairi, 1988; LaSalle & Conture, 1995) disfluencies appear to be quite common in the speech of children who stutter, and may help differentiate children who stutter from children who do not stutter (LaSalle & Conture, 1995). Thus, it would seem necessary for clinicians to account for the occurrence of these types of disfluencies when analyzing a client’s speech fluency.

**Counting stuttering.** Even though the classification schemes described above are based on empirical research regarding the disfluencies exhibited by individuals who stutter, the basic fact remains that not all “within-word” or “stuttering-like” or “less typical” disfluencies are actual instances of stuttering. Thus, many clinicians and researchers prefer to count instances of stuttering. In general, the
speech behaviors judged to be stuttering will be similar to those labeled as “within-word” or “stuttering-like,” or “less typical;” however, counting stuttering can allow more flexibility in judging speech behaviors than the behavioral guidelines discussed above. For example, a brief, non-tense prolongation may not be perceived as stuttered, so it would not be counted as stuttered even though, technically, it is a within-word or stutter-like disfluency.

Of course, counting instances of stuttering is not without problems of its own, because different clinicians, and different clinics, may have vastly different criteria for determining what constitutes stuttering (e.g., Cordes & Ingham, 1995a; Ham, 1989; Ingham & Cordes, 1992; Kully & Boberg, 1988). In fact, depending on how the term “stuttering” is defined, it may be quite impossible for a clinician to ever know that stuttering has actually occurred. Perkins (1990; Perkins, Kent, & Curlee, 1991) defined stuttering as a “loss of control” experienced by the speaker, regardless of the nature of the speech output perceived by the clinician. Thus, in order to know with certainty that stuttering has occurred, a clinician must check with the client immediately after the production of the disfluency (Moore & Perkins, 1990). Therefore, problems associated with measuring various types of disfluency behaviors are not solved simply by choosing to count instances of stuttering.

So what should I count? In sum, there are problems with counting either disfluencies or stuttering, but this should not deter clinicians from making such measures. Based on the preceding discussion, it would seem beneficial for clinicians to measure a variety of disfluency types because this provides useful diagnostic information. Still, categorically defining all “within-word” or “less typical” disfluencies as instances of stuttering is problematic. Thus, it would also seem appropriate for clinicians to specifically describe which of their client’s disfluency behaviors were judged to be stuttering, as well as the criteria for making such judgments. Such an approach would combine the diagnostic strengths inherent in measuring a broad range of disfluency types with the flexibility and clinical relevance of measuring instances of stuttering, while minimizing problems associated with following either approach too rigidly.

Counting Words or Syllables

A brief review of the literature on stuttering measurement reveals that some clinical researchers prefer to count the number of syllables produced in a given speech sample (e.g., Boberg & Kully, 1985; Campbell & Hill, 1987; Costello & Ingham, 1984; Riley, 1994), whereas others count the number of words (e.g., Conture, 1990a; Conture & Caruso, 1987; Riley, 1980). In general, the differences between these two units of measurement appear to be relatively minor because it is possible to convert from one to the other by simply multiplying by a constant representing the ratio of the number of syllables per word (Andrews & Ingham, 1971; Conture, 1990b; Conture & Caruso, 1987; Culatta & Goldberg, 1995; Ham, 1986). For young children, the number of words in a sample will be quite similar to the number of syllables (i.e., the constant will be close to 1.0) because many of the words produced by young children are monosyllabic. For adults, the correlation between number of syllables and number of words will not be as strong, though a conversion is still possible using a constant that more accurately reflects the average number of syllables per word (e.g., 1.5, see Andrews & Ingham, 1971).

For some researchers and clinicians, there may be theoretical or practical reasons to prefer one unit over the other. For example, the syllable is presumed by many to be the basic gestural unit in speech production, whereas the word is presumed to be a basic linguistic unit in speech formulation. Also, depending on the specific measurement technique selected (more on this below), there may be some benefit to selecting words or syllables as the basic counting unit. For example, if a clinician is working from a verbatim transcript, it may be easier to count the number of words produced than it would be to sound out each word in order to count the number of syllables. If, on the other hand, the clinician is using an on-line technique based on listening to the client while he or she is talking, then counting syllables may be easier because the clinician can more easily listen to stressed peaks in the speech flow. Thus, a clinician should select whichever unit is appropriate for his or her overall measurement technique. Regardless of what clinicians select, it would seem appropriate for clinicians to be consistent in the unit they use with their clients, and they should specify their unit of measurement when communicating results to colleagues or clients.

Event-Based or Time-Interval-Based

Based on concerns regarding the reliability of various techniques for counting instances of stuttering or disfluency (discussed in greater detail below), another basis for measuring stuttering behaviors has recently been proposed (Cordes & Ingham, 1994b, 1994c, 1995a, 1996b; Cordes, Ingham, Frank, & Ingham, 1992; Ingham, Cordes, & Finn, 1993; Ingham, Cordes, & Gow, 1993). This technique is based not on counting instances of disfluencies or stutterings (event-based measurement), but on determining whether stuttering occurs during a given time interval. Thus, rather than listening to a 100-word conversational speech sample and counting the number of disfluencies (or stutterings) that occur, the clinician would divide that sample into a given number of intervals lasting, say, 4 or 5 seconds each (so a 1-minute speech sample would contain 15 4-second intervals) and then determine how many of those intervals contained disfluencies or stuttering.

Early reports on this technique (Cordes et al., 1992) suggested that time-interval measures suffered from the same poor inter-judge agreement (i.e., agreement between judges) and intrajudge agreement (i.e., agreement of a

---

6 Other clinicians have argued that it is possible for a sensitive clinician to learn to identify when their clients are experiencing this loss of control (e.g., Manning, 1996).
单个法官（或本身）之前被识别为基于事件的测量（协议指数低于60%；例如，Curlee, 1981; MacDonald & Martin, 1973; Young, 1975）。更多的研究显示，这种协议是可能的，并且可以显著提高并内和内内判断协议（协议指数在大约85%）通过培训判断者，他们的工作成果被发表于Cordes & Ingham, 1994c, 1996b; Ingham, Cordes, & Gow, 1993; see also Costello & Hurst, 1981)。在我们的研究中，这些研究发现，为了提高判断协议，可以被从个体与更高的内内判断协议 (Ingham, Cordes, & Finn, 1993) 的优点。

虽然时间间隔测量技术具有的显而易见的优缺点，即这种技术也有可能使在临床实践中，它不能准确地说明这种技术的可能应用。多的关于时间间隔的测量技术在随机化顺序数据生成计算机控制的磁盘。尽管这种技术显示更高的一致性，它不是为了被转移给临床环境的。第二，时间间隔的变化可能会影响判断的特性，即在给定的样本中，如样本的言语样本可能会产生不同的结果，取决于它是否被分析为3秒或4秒的间隔 (Cordes & Ingham, 1994b)。

第三，判断的间隔时间不会影响允许研究者和临床医生对言语和语言生产的特征的特定位置的异常。

最后，也许最重要的是，它不是非常清楚，在定义的临床决策过程。换句话说，它不清晰的定义数据表明，每个被选中的人，如三个语音声段，提供了相同的临床信息，如数据，表明每个被选中的人，三次听觉声段在100个语音声段中。这是非常重要的，因为不同的语音声段时间间隔可能会产生不同的语言声段，取决于对相关的行为和情况。因此，进一步的研究在听觉的使用和应用性方面，这项技术的完成，它似乎合理，至少，对临床医生的持续计数，将有助于确保可靠性，而工作努力来确保他们的措施。

**DIFERENT TECHNIQUES FOR MEASURING STUTTERING**

在确定哪种行为应该被测量，它也非常重要的是考虑，这些行为将被测量。虽然这个主题没有被如此关注，在研究文献中，很明显，有更多的不同的技术可用于测量结结行为。

**On-Line or Off-Line Measurement**

一个早期的决定，一个医生必须做出的是收集数据“on-line,”（或“实时,”，即，当患者正在说的话）或“off-line”（即，从音频或录像带之后的会议是关于）。在不同技术的收集有频率数据，通常是言语处理上的治疗会议。然而，那里有几种潜在的缺点，需要在收集数据前应记录为“on-line” (Boberg & Kully, 1985; Conture, 1990b; Ingham, 1993; Neilson & Andrews, 1993; Ryan, 1979)，它似乎不可能收集到详细的数据，如一个概括的脚本（特别是当临床医生参与对话）。因此，医生可能需要在收集语音样本之后的时间，对决定医学行为或再分析的言语样本是更困难的。

最终，而且最重要的是，数据关于收集on-line技术的可靠性不是目前可用的，临床医生必须小心，在收集on-line数据时，他们的言语行为。虽然有足够的时间和确定可靠性，它似乎在控制on-line测量是适当的，对于评估某些结结行为。

一种方法是，on-line分析涉及计算机化的计数系统（例如，Boberg & Kully, 1985），在其中，医生按下按钮的一个键的每个词（或声音）产生的，由一个按钮，为每个有困难的言语（或结结）提供的。然后，计算机可以自动计算频率的困难（以及说的速率）。一个缺点是，它们无法提供关于困难类型的不可用的信息，尽管它们似乎有可能，对一个系统是开发的。

另一种方法是在收集on-line分析涉及收集困难的计数形式，如由Conture (1990b, Conture & Yaruss, 1993) 等人（例如，Culatta & Goldberg, 1995; Riley, 1981, 1994) 所描述的。这些形式通常提供了一系列的空白空间分成块的100。医生然后在与一个 difficulty words with a dash and difficulties with an “X” or an abbreviation indicating the type of difficulty produced (e.g., SSR for sound/syllable repetition). This allows the clinician to quickly calculate the frequency of difficulties, as well as the relative frequency of different difficulty types (see Conture, 1990b).7

**Disfluency-Based or Transcript-Based Techniques**

当收集off-line分析，临床医生有额外的选项，一个概括的脚本，而不是

---

7 Similar forms can be used for on-line assessment of the average duration of disfluencies, the client’s average articulatory speaking rate, or other behaviors of interest.
than just tallying the number of disfluencies produced during a speaking sample. A verbatim transcript has the benefit of providing a considerable amount of data beyond frequency counts. A number of techniques for preparing transcripts has been described (e.g., Campbell & Hill, 1987; Rustin, Botterill, & Kelman, 1996). In one such technique, Systematic Disfluency Analysis (SDA, Campbell & Hill, 1987), transcripts of 200 syllables of conveyed message are prepared from audio/videotaped speech samples. Speech disfluencies are identified using the broad range of “more typical” and “less typical” disfluency types summarized in Table 3, including any instances of “multicomponent” or “clustered” speech disfluencies. Next, various durational (e.g., number of iterations per repetition, length of prolongations), audible (e.g., increases in loudness, pitch changes), and visible (e.g., nonspeech behaviors, physical tension) aspects of disfluencies are noted on the transcript. Finally, each disfluency is scored on a weighted point system, and a severity index is calculated (Campbell, Hill, & Driscoll, 1991). By providing information about a broad range of disfluency types and accompanying qualitative features within the context of a speaker’s message, a transcript-based method such as the SDA can give the clinician access to considerably more information for making clinical decisions than frequency counts, and may be particularly useful at the time of an initial diagnostic evaluation or at important “transition points” in treatment.

The most apparent drawback to such detailed techniques is the time required to prepare the transcript. As with many other endeavors, however, there is an expected trade-off between the amount of time devoted to completing the analysis and the amount of information obtained from the analysis. A clinician who completes a verbatim transcript such as the SDA can assess the relationship between linguistic factors and the occurrence of disfluencies in a way that is not possible with a technique that simply tallies the number of words (or syllables) and disfluencies (or stutterings). Unfortunately, as with the on-line measurement techniques described above, the reliability of detailed analysis techniques such as the SDA has not yet been empirically demonstrated. Therefore, although there is a tremendous amount of additional information to be gained, clinicians may still need to be cautious regarding their use of detailed transcription systems until further research on the reliability of these techniques has been conducted.

So What Measurement Technique Should I Use?

Again, the final determination of which measurement technique should be selected is dependent on whether the measurements are being conducted during a diagnostic evaluation or during a treatment session. In the end, clinicians and their clients will probably be well-served by a combination of both on-line and off-line measurements, as well as a combination of both disfluency-based and transcript-based analysis techniques. At the time of an initial diagnostic evaluation, an on-line frequency count while the client is talking to a peer or a parent can provide a fast, quantitative indication of the client’s degree of (dis)fluency in that situation, while still allowing the clinician to present results to the parents or client immediately following the evaluation. Following the evaluation, the clinician can review the audio- or videotapes off-line, either to refine the original on-line measures, to create a verbatim transcript to assess how other aspects of speech and language might relate to the client’s stuttering, or to help with planning treatment. Later, during treatment, the clinician can make regular and consistent measurements of certain treatment variables during each session (e.g., quick on-line tallies of the frequency and types of disfluencies or the use of modification techniques), combined with less-frequent, but more detailed measurements of other important variables (e.g., off-line measures of the relationship between situational factors and stuttering collected at the beginning or end of a school quarter). Finally, at the conclusion of treatment, the clinician can again conduct a more detailed transcript-based analysis to document changes in the client’s fluency from the beginning to the end of treatment. In this way, clinicians can obtain relevant and important data concerning their client’s stuttering behaviors without spending all their time doing assessment.

RELIABILITY OF STUTTERING MEASUREMENTS

One important issue that has been raised a number of times in the preceding discussion is the reliability of various measurements techniques. Several authors (Cordes, 1994; Cordes & Ingham, 1994a; Curlee, 1981; Kuly & Boberg, 1988; MacDonald & Martin, 1973; Onslow et al., 1992; Tuthill, 1946; Young, 1975, 1984) have discussed a number of problems with clinicians’ ability to make reliable judgments of stuttering behaviors. In general, early reviews suggested that it was possible to achieve high levels of agreement for overall stuttering counts, but not for individual instances of stuttering (e.g., Young, 1984). In a more recent analysis, however, Cordes and Ingham (1994a) concluded that measurement reliability is problematic across the whole range of stuttering research and highlighted the potential benefits of using interval-based measures for improving measurement reliability (although see the concerns raised above).

Findings regarding low measurement reliability appear to be somewhat at odds with numerous empirical studies of stuttering in children and adults in which researchers routinely report acceptably high levels of inter- and intrajudge measurement reliability. This discrepancy may be due, in part, to the fact that the statistical techniques used to calculate measurement reliability in many studies have not always been the most rigorous (Lewis, 1994). It is also likely that researchers obtain higher agreement because of their extensive training and the preparations they make before beginning a particular study (Cordes & Ingham,
WHERE DO WE GO FROM HERE?

Based on all the concerns presented in the literature, some clinicians may be left with the feeling that it is essentially impossible, given our current level of understanding, to collect meaningful, reliable, and clinically relevant measures of stuttering behaviors. It is this author’s opinion that this would be an unduly pessimistic reading of the literature. Given the importance of stuttering measures in the diagnosis and treatment of individuals who stutter, we must resist any tendency toward “paralysis by analysis,” and not allow confusion or conflicting philosophies concerning measurement techniques to prevent us from making the most accurate and reliable measures possible.

The Role of Clinical Training

In order to improve clinicians’ ability to make accurate and reliable measures of clients’ stuttering behaviors, it is important that training centers ensure that the techniques they teach are reliable, and that students are given adequate time to develop their clinical skills. The full responsibility for this cannot fall only on training centers, however. Clinicians, too, must take responsibility for making appropriate decisions concerning the measures they make. And, they must commit to practicing the measurement techniques until they are able to make accurate and reliable measures. One way to accomplish this outside of a training institution is for clinicians to make several video-recordings of their own clients, then score and re-score the speech samples, using both on-line and transcript-based techniques, until they are able to achieve a high degree of intra-rater reliability. Clinicians should also consult with one another about the measures they make in order to improve inter-rater reliability.

A Role for Stuttering Specialists?

The advent of specialists in stuttering diagnosis and treatment may provide an additional means for addressing the problems associated with measurement reliability. Specifically, stuttering specialists will likely have completed additional practice and training in the identification and measurement of stuttering behaviors. Therefore, in addition to providing consultations regarding the diagnosis and treatment of individuals who stutter, stuttering specialists should also prove to be a valuable resource for clinicians seeking to improve their ability to measure their clients’ stuttering behaviors.

CONCLUSION

It is true that there are many different opinions about how best to measure stuttering behaviors. It is also true, however, that clinicians have been meaningfully assessing and treating stuttering for decades. Therefore, rather than causing us to stop measuring clients’ stuttering behaviors,
the concerns raised in the literature should provide an impetus for further clinical research designed to address some of these important issues, for example:

- How can we improve the reliability of stuttering measurements in clinically relevant ways?
- How do the judgments of whether or not a child is at risk for stuttering made in one clinic compare to those made in another clinic?
- Is there a clinically meaningful difference between measuring stuttering in terms of words or syllables?
- How can we improve the training of student-clinicians to improve the reliability of their measures of stuttering behaviors?

Until answers to such questions are available, it will be impossible to know with absolute certainty that the measures we make are the “right” ones. Meanwhile, however, clinicians should continue to do their best to make meaningful and justifiable measures of their clients’ stuttering behaviors, while remaining aware of the potential pitfalls associated with the various measurement techniques they select.

ACKNOWLEDGMENTS

The author is grateful to Anthony Caruso, Walter Manning, and Patricia Zebrowski, as well as to Diane Hill and June Haele Campbell for helpful comments on an earlier draft of this manuscript.

REFERENCES


