Stem-, Spraak- en Taalpathologie
1999, Vol. 8, No.4, pp. 220-231

32.8310/99/0804-220

Stuttering severity measurement through objective and perceptual event-related methodologies, and interval scoring

K. Bakker

Southwest Missouri State University, Springfield, USA

The long lasting practice of stuttering severity measurement through objective and/or perceptual event-related procedures has been challenged in recent research literature. Interval scoring has been suggested as an alternative procedure because of its empirically demonstrated improvement of intra- and inter-rater reliability compared to the event-related procedures. Yet, this advantage has not yet tempted most practicing clinicians. Moreover, a number of questions have surfaced regarding possible limitations in validity of the interval based strategy for stuttering severity measurements.
While interval scoring improves aspects of reliability in stuttering evaluations, and to a significant degree, it's ability to effectively depict the nature, and the 'subjective' components, of stuttering severity has not been properly tested, and, unfortunately, while such tests may not be logically possible. This paper reviews and evaluates advantages and disadvantages of traditional and current event- and interval based procedures for the measurement of stuttering severity. In doing so it addresses (1) how each measure works, (2) what types of results are available with it, and (3) how differences in such results can differentially affect clinical decisions.

Introduction

Usually during fluency evaluations, following the diagnostic determination that one's client demonstrates the clinical features of stuttering, measurement of its severity is next in line. It is easily defended that the objective of any stuttering treatment should be a clinically meaningful reduction of stuttering severity. Importantly, stuttering severity measures may be frequently needed throughout therapy (e. g., for determining the client's baseline performance, monitoring therapeutic progress, and, ultimately, forming decisions about discharge from therapy). Given the obvious importance of stuttering severity measures one would expect that agreement on *what* should be measured and *how* should be rule rather than exception. Unfortunately, as will be seen on the following pages, the opposite appears to be the case at present.

Correspondentieadres: K. Bakker, Southwest Missouri State University, Springfield, MO 65804-0095, USA.

This paper was developed in response to the following needs: (1) to establish a *scope* for quantifiable measures of stuttering severity (e. g., what should be measured?), (2) to determine which aspects of severity are reasonably measurable, modifiable, and meet commonly accepted standards for validity and reliability for reflecting their targeted characteristics, and (3) to determine which of them are *feasible* for routine clinical application.

Discussion of the event-related stuttering measures

Stuttering frequency measures

The most common solution for the measurement of stuttering severity involves some representation of *frequency* of its *occurrence.* Usually, this aspect is made proportional to the overall range of opportunities stuttering can manifest itself. There are two possible ways to do this. That is, stuttering can be made proportional to the amount of speech produced (e.g., expressed as a percentage of syllables stuttered, or %SS; see for example Costello and Ingham, 1984). Alternately, a number of stutterings can be made proportional to overall duration of speech in a sample (e.g., expressed as number of stutterings per minute).

The 'percentage of syllables stuttered' measure is considered representative across a wide range of speaking styles and conditions. There is an increasing preference for expressing frequency relative to the numbers of syllables, rather than words, produced by the client. There are a number of reasons for this preference. Syllables are less variable in length and duration than are words. Also, syllables reflect the physiological and kinematic adjustments more directly as "words" as a unit of sample size tend to be more abstract in nature. Moreover, syllables are fairly discrete perceptual events that match, but barely so, the limits of human sensory motor reaction capabilities. Finally, the choice of the syllable metric is attractive, because a number of commonly used definitions of stuttering define dysfluency as a feature that is confined to syllable boundaries. This, then, suggests that it would be appropriate for clinicians to search and sort among all spoken syllables for those that are, or are not, stuttered.

Methodological questions about frequency measures:

1. Are stutterings really confined to syllable boundaries?
   Some have expressed doubt that stuttering is necessarily confined to syllable boundaries (Curlee, 1981). If one assumes, for example, that stuttering represents instances during which a perceptual threshold (Martin and Haroldson, 1981) is exceeded, there is no need to link stuttering to specific units, such as syllables or words. Stuttering, according to this hypothesis, could well follow a waxing or waning pattern and, thus, exceed syllable boundaries. Such a pattern, among others, would be consistent with some physiological and acoustic observations that have suggested that perceptually fluent speech of stutterers differs from that of normally fluent speakers. To the extent that physio-acoustic differences represent minimal forms of stuttering, the term *% of syllables stuttered*, is a misnomer. That is, the

terminology assumes that syllables *can* be discretely classified as those that are, and those that are not stuttered. It would seem preferable to change the name of the measure in order to account for a greater variety of interpretations of stuttering. The stuttering frequency measure, perhaps, should rather be referred to as the 'number of stutterings per 100 syllables'.

2. Are all stutterings stuttering to the same extent?
   Frequency measures are insensitive to many of the unique differences that exist between individual instances of stuttering. For example, the potential effects on severity of such aspects as stuttering duration, effort, and perceived abnormality are *not* reflected by such measures. Another limitation is that stuttering frequency measures are independent of speech rate. They do not reflect whether a stuttering frequency value represents speech that was slow, normal, or relatively fast. Because of these potentially confounding factors, frequency measures can only be expected to reflect stuttering severity in part and should be accompanied, if possible, by other measures, or descriptors, in order to provide a complete account of the severity of an existing fluency problem.
3. Are stuttering frequency measures clinically feasible?
   The # of stutterings per 100 syllables measure is not popular with clinicians who are in need to use it frequently, often on-line, and for sustained periods of time. This, of course, is because of the need to count all syllables and stutterings in recorded speech samples. It is a notoriously demanding task, especially during normal and near normal rates of speech delivery, unless transcriptions can be used to determine the syllable counts. But importantly, the typical clinician's ability to reliably identify all of the syllables of a running speech sample, has never been empirically assessed.

   The "# of stutterings per 100 syllables" measure, moreover, looses its effectiveness for differentiating clinical progress when stuttering is infrequent, or reduces in frequency during later stages in therapy. Under these circumstances it is more effective to express stuttering frequency in the form of "average length" of periods of stutter free speech expressed in # of syllables.

An alternate way for expressing stuttering frequency is to measure *stuttering rate*, or the average number of stutterings per minute of speech production. This measure has as the obvious advantage in that it does not require a clinician to count all of the syllables spoken. Nevertheless, this advantage represents a possible weakness as well. As numbers of stutterings are not directly made proportional to the amount of speech produced, stuttering rate is potentially confounded by variations in speech rate. This problem may be avoided by accompanying this measure with information regarding the client's rate, which requires that all syllables are counted after all. The convenience of the stuttering rate measure, then, is confined to those situations where clinicians believe that it is unlikely that the measure is confounded by speech rate variations.

Aside from the possible speech rate biases, stuttering rate measures may be also biased by any pauses that are disproportionately long in duration. It is advisable that the timing device used is interrupted for lengthy breaks in speech production. Unfortunately, there are no easy ways to standardize pause omissions from the stuttering rate me-

asures, let alone when administered manually, and on-line. This makes the stuttering rate measure to some extent arbitrary, especially in spontaneous speech where pauses, and periods of silence, occur to a greater extent then, for example, during oral reading (Bakker and Gregg, 1994).

Finally, as was the case with the # of stutterings per 100 syllables measure, the stuttering rate measure looses its effectiveness when stuttering occurs infrequently. In this event, stuttering rate is better expressed as the average duration of perceptually stutter free periods expressed in seconds, or minutes.

Stuttering duration-related measures

Frequency measures do not cover all of stuttering severity. Nevertheless, frequency measures are the only measures that are considered when reporting levels and changes in levels of stuttering severity. Whether or not this is intentional, to limit severity evaluations only to the frequency aspect implies that this aspect is considered the primary problem, and, also, that a reduction in frequency is considered the ultimate sign of therapeutic progress. It is reasonable to expect, then, that stuttering frequency is all that changes in response to such therapies.

Where frequency measures represent only one aspect of stuttering severity, *stuttering duration-related* measures could cover much of the remaining ground. Occasionally, a client demonstrates a low number of stutterings, but these blocks are abnormal in appearance and long in duration. Clearly, for this type of client stuttering frequency alone would lead to a misleading representation. It does not reflect, for example, what it is that is severe or abnormal about the speech of the particular client. Severity, in this case, would be more effectively expressed when frequency and duration based measures are reported together, permitting that they can also be weighed against each other. There are a number of compelling reasons for including durational aspects of stuttering in the equation of severity. It has been reported, for example, that early treatment gains tend to be reflected in stuttering durations rather than their frequencies (Conture and Caruso, 1987). Therapeutic progress, in a durational sense, would not be detected by frequency measures alone. Moreover, by keeping separate records on stuttering frequency and duration, one keeps the option open to determine to which extent a stutterer's fluency problem is based on either aspect, and develop individualized clinical goals that reflect this distinction.

It is reasonable to assume that frequency and duration based measures have independent contributions to the overall severity of a fluency problem. Some evidence for this relationship exists. Low to non-existent correlations between stuttering frequency and duration, among others, were observed at three previous occasions (Bloodstein, 1944; Johnson and Colley, 1945; Zebrowski, 1994). Bloodstein (1995), who conducted one of these studies considered the weakness of this relationship a sign of the apparent lack of criterion-related validity for using durational measures to represent severity. Along a different vein, this author proposes that durational measures of stuttering apparently reflect aspects of severity that are not covered by frequency measures. This, then, would imply that durational measures are unique and probably necessary in certain assessments. The ultimate validity test of durational measures should come

from studies that determine the relationship between stuttering duration and judged severity of individual instances of stuttered speech determined by experienced clinicians. Such evaluations, to the knowledge of this author, have yet to occur. But unfortunately, such research could be compromised by recent findings that such expert clinicians may not meaningfully agree on the identification of individual stutterings, let alone determine their durations. Nevertheless, such evidence could play a factor in appraising the need for durational measures in describing aspects of stuttering severity.

There are inherent difficulties in measuring the durations of individual stuttering moments. Assuming that stutterings are specific events that have discrete and perceptual beginnings and ends, durations are still difficult to measure accurately because of their inherent time frame. Stutterings have been estimated, or judged, to be around 1 second in duration, while the durational distribution is positively skewed. This suggests that more than half of the stutterings are less than a second while only a minority of stutterings is disproportionately long. As a result, most stutterings have durational properties that do not permit them to be determined with great precision. This is probably why average stuttering durations, such as they may be determined with a stopwatch, have never caught on as a practical or feasible measure for stuttering severity.

Each individual stuttering duration requires two separate sensory-motor reaction times (and subsequent decision times), which under optimal conditions, are about 200 ms each. One reaction time involves onset of stuttering, while a second one involves its termination. An undetermined amount of extra time is needed, however, when stutterings are of an ambiguous perceptual nature, and are either very brief in duration, or behaviorally complex. While the onset and termination-related reaction time delays are in the same direction, and possibly cancel each other out, the possible variations in stuttering identification judgment times could make durational measures, to say the least, a suspect representation of stuttering severity.

It is reasonable to hypothesize that the identification of stuttering is more reliable and accurate when stuttering moments are relatively long in duration (e.g., exceed one second in duration), while, also, the impact of human reaction- and judgment-times is insignificant in these cases. This then suggests the possibility that, to the extent that durational aspects of stuttering are considered desirable parameters for reflecting stuttering severity, only long instances have satisfactory metric properties

Currently popular measures of stuttering duration are (1) the average three longest moments of stuttering (Riley, 1972, 1994; Costello and Ingham, 1984) and (2) the estimate of the 'typical stuttering duration' which, for example, can be based on the mean of ten representative stuttering durations (e.g., Conture and Caruso, 1987). Both measures in combination are suggestive for the amount of spread in duration across entire speaking samples, but it is necessary that a sample is recorded on video, and replayed, multiple times in order to locate a sufficient number of typical, or the three longest, instances of stuttering.

The mean of the three longest stutterings measure has now become a popular way of expressing a client's worst scenario case of stuttering duration. The popularity of this measure is easily understood. Long stutterings do not suffer nearly as much from the methodological problems associated with measuring brief and fleeting moments

of stuttering. Despite this, not much is currently known about the metric properties of the maximum duration measure. We do not know, for example, to what extent it can be considered a stable measure and how long a sample needs to be before this measure is representative for the client's extreme durations. In selecting a maximum, rather than typical duration of stuttering, furthermore, a potential methodological bias is created, which, to this author's knowledge, has not been addressed in the literature. That is, with incrementally longer samples, the theoretical chance that still longer stutterings may occur increases as well. When stuttering is also relatively infrequent, low numbers of stutterings invalidate this measure as well. For example, there have to be at least three stutterings to measure it, and in that situation the resulting value would be identical to the mean stuttering duration. The three longest stutterings measure, then, is confounded by variations in duration and length of speech samples, as well as the numbers of stutterings demonstrated during these periods.

The aforementioned problems could be partially fixed by adopting a standardized sample size for the determination of the mean three longest stuttering blocks measure. Perhaps the most convenient solution would be to apply the three longest stutterings measure to only a fixed portion, for example the first one hundred syllables, of a speech sample, while also a minimum number of stuttering moments can be set before the measure is reported. Clearly, research is needed to further explore the metric properties of the three longest stutterings measure, its stability, and the conditions under which its validity is potentially compromised.

Other than focusing on the most extreme, or the most representative durations, one has the option to compute the mean of all stuttering durations. Obviously, when conducted by hand, this is impractical. Nevertheless, it is a feasible option when computerized recording procedures are used. Computers, that is, present the option of storing and retaining all individual stuttering durations marked by an observer, and consequently compute a range of statistics that apply to these data. Unfortunately, as most of the stutterings are very brief and fleeting in duration, one might rightfully question the precision and reliability of a measure that so heavily taxes one's sensory motor reaction and judgment capacities. This measure may not be an option in on-line manual determinations. At any rate, there is little empirical evidence that relates to these important measurement issues.

Some of the demands associated with durational measurements of stuttering dysfluency can be circumvented by computing another, derived duration related, measure. Rather than relying on the actual durational values themselves, one could express the percentage of stutterings that exceed 1 second in duration. This measure has an intuitive advantage in that it expresses what proportion of the client's stutterings exceed the average stuttering duration in the population. Also, the suggested cut off point of one second would help eliminate possible confounding created by differential perceptuo-motor and judgment related abilities of clinicians. Although the measure is impractical when conducted manually with a stopwatch, it is very feasible with computerized scoring procedures.

While reducing the impact of human error in measuring duration, the proportional durational measure still depends on the correct identification of **all** instances of stutte-

ring and, thus, the observer's correctness in identifying even the briefest instances of dysfluency. There are no empirical data, at this time, that reflect on the potential usefulness of the proportional durational measure in addition to any of the aforementioned duration related measures.

Finally, rather than focusing on durational measurements of individual occurrences of stuttering, there is the option to look at the durational aspect in a general sense. One could compute the percentage of articulatory time that a client is engaged in stuttering (Starkweather, 1993). By choosing a general rather than event specific durational measure, it is possible that the influence of reaction time and judgment related errors average out in the long run. Also, the effect of differential judgments regarding identifications of even the briefest instances of stuttered speech, could "average out" with this procedure. To convert the stuttering time measure into a percentage of articulatory time, furthermore, is a highly desirable clinical measure because it allows the impact of stuttering frequency and duration to be directly compared. In other words, clients can be characterized as having either a frequency-, or duration-, specific type of severity. Yet, determination of the percentage of speaking time stuttered is very new, and has not often been tested under controlled circumstances.

Measures that reflect the incidence of dysfluency types

The most methodologically explicit approach to stuttering severity is to operationally define the behaviors that are considered stuttering. Following this approach, only the evidence of specific types of dysfluencies, usually defined as audible or silent part word repetitions and prolongations, are considered in measurements of stuttering severity. On paper, this approach is very straightforward and would be expected to lead to reliable descriptions of severity. After all, all of research methodology recommends the use of operational definition in order to make experimental treatments and measurements replicable. To the extent that experiments can be considered measurements of the truth, operational definitions are used to enhance reliability of the chosen methodology.

Unfortunately, in the case of stuttering use of explicit behavioral definitions has not produced the expected advantages in comparison to measurements that were based on the mere perception that stuttering occurred. Even though clinically significant reliability between and among raters is observed when total numbers of behaviors are assessed, this was not the case when reliability concerned specific individual instances of stuttering. In the end perceptually scored stuttering too failed to meet generally accepted clinical norms for reliability and did not produce any advantages over behaviorally explicit stuttering observations.The fact that reliability among and across observers does not have any benefit from the use of operational definitions is a deviation from the norm, and, thus, requires explanation. Of course, it may be, as some have suggested that the use of operational definitions has not been properly tested (Hegde, 1995). After all, a number of the crucial comparisons between perceptual and behaviorally defined observations of stuttering have involved naive, or minimally trained subjects. It is possible, also, that there is something wrong with the operational definitions of stuttering that are routinely used. However, it is more likely that operational

definitions were not helpful because stuttering itself is unreliable in appearance. It presents itself through many, often brief, moments of dysfluency forcing classifications to be made in borderline perceptual conditions. A number of the stuttering behaviors, also, are complex and compounded and, perhaps, defy any known classification. After all, even when depending on behaviorally explicit analysis schemes, stuttering identification is still largely a perceptual task.When evaluating the reliability of the identification and measurement of stuttering there is a tendency to not differentiate the characteristics of individual stuttering moments. This is unfortunate, as this global approach possibly hides many of the specific reasons for why stuttering related judgments differ between raters. It may well be that when judgments are confined to only the most severe instances of stuttered speech, point by point reliability checks in highly. For example, if one is to judge blocks that are at least several seconds in duration and evidence much effort, there may not be any differences in the judgments among and between observers.

It may well be that reliability of stuttering identification is low only because of the fact that most stuttering moments are difficult to perceive. To the knowledge of this author no studies have treated *reliability as a variable by itself*, and made it dependent on specific conditions under which perceptions of stuttering occur. It is well possible that under less constraining conditions both perceptual and behaviorally explicit approaches produce satisfactory levels of intra- and inter-observer reliability.

There are two opposing positions regarding the need for behavioral descriptions in the identification and measurement of stuttering. On the one hand there are those who believe that stuttering isn't stuttering until it is perceived as such. Taken to its extreme this position defines stuttering entirely as perceptual in nature, thus reducing any need for objective or quantifiable behavioral descriptors. At the other extreme is the belief that perception alone is too subjective to be acceptable for the identification of stuttering behaviors, because it is not made explicit which behavioral features were responsible for these perceptions. Taken to *its* extreme this position defines stuttering as a phenomenon in the objective reality thus equating the behavioral definitions with the stuttering itself. Both extremes have their own unique problems with validity, and thus reality. Where topographically oriented analysis schemes are inaccurate, or incomplete, in depicting the core features of stuttering, the 'perception only' position is vulnerable because it defies objective checks of validity, and depends on comparing and analyzing the perceptual processes involved in stuttering identification.

A middle of the road position seems possible. This position considers stuttering as a phenomenon in the objective reality that acquires its status of stuttering only after it is perceived as such (Young, 1984). In this sense, behavioral descriptors serve only as guidelines, but do not enforce which events should be considered instances of stuttered speech. According to this position, it may well be that a number instances, characterized by the presence of part word repetitions, or prolongation, are not considered stuttering because they were not perceived as such. In contrast, a number behaviorally compound instances (that is, mixtures of segmental repetition and prolongation behaviors) could be accepted as instances of stuttering despite the fact that they did not fit any accepted operational definition.

If one assumes the need for behavioral definitions of stuttering one also carries the burden of demonstrating the validity of these descriptions. This is not an easy matter. There is little direct support from research, or theoretical development, that can be of help in defining the unique and essential features of stuttering. The best evidence for the currently most popular operational definitions of stuttering comes from two sources. One form of evidence is the observation that stuttering type dysfluencies, that is audible or silent PWRs and PRLs, are among the behaviors uniquely displayed by those who stutter (e. g., Johnson, 1959). Another source of evidence, suggestive for the uniqueness of certain types of stuttering behaviors, involved experimental manipulations of their incidence. Among the findings that resulted was the observation that certain dysfluencies are more likely to be perceived as stuttering than others, given equal exposure, while it was also true that some dysfluencies became perceived as stuttering when they occurred frequently enough. These findings, importantly, support both for the position that it is necessary to distinguish objective types of dysfluency as well as the position that stuttering is perceptual in nature.

It does not seem possible at this time, without question, to specify all of the unique aspects of the symptomatology of stuttering. Yet, it is clear that, in the end, stuttering represents features both in the perceptual and physio-acoustic domains. Perceptual judgments and observations of stuttering should, ultimately, be traceable to the physio-acoustic reality. If not, stuttering would surprising well fit the definition of "delusion" such as this term is used in psychiatry. It is probable that traditional operational definitions of stuttering represent the majority of what is perceptually identified by others. Without necessary background knowledge, consensus, and believable theoretical support, perhaps the only certainty about stutterings is that a client experiences them, and that they are perceived as stuttered by certain others. The question of validity, of what should be measured to reflect stuttering severity, then, would reside at least in part with the owners of these experiences.

Interval based stuttering severity measures

As the previous pages have illustrated, stuttered speech is traditionally measured in terms of counting numbers of perceived moments of stuttering (Martin & Haroldson, 1981) or in terms of counting numbers of certain types of disfluency (e. g., Johnson, 1959). One well-known problem with these systems, however, is that the resulting data are not replicable; that is, both inter-judge agreement and intra-judge stability are known to be problematic for observer judgments of stuttered speech (Cordes & Ingham, 1994a; Ham, 1989; Kully & Boberg, 1988; Onslow, Gardner, Bryant, Stuckings, & Knight, 1992).

A potential solution to the reliability problem with individual event-related judgments, is the use of interval-based measurement procedures. Interval measurement, to be sure, is not new and has a demonstrated record as a standard behavioral measurement option in behavioral sciences (Kazdin, 1982). It is recommended for use when the reliability of event-based measures is potentially problematic. It's use with

the measurement of stuttered speech is frankly not entirely new (e.g., Shaw & Shrum, 1972). Interval-based measurement systems for stuttering, however, have been developed systematically in a series of recent papers by Ingham, Cordes, and colleagues since approximately 1990. The results of interval based judgments in experimental settings have demonstrated repeatedly, and without exception so far, that intra-judge and inter-judge agreement levels are higher for interval-based measurements of stuttering than for event-based judgment procedures.

The interval-based stuttering severity determination procedure is a relatively new option for clinicians. When supported by appropriate computer software, its implementation is easy and almost transparent to its user as well as the client. That is, the amount of effort by clinician is minimal compared to identification, judgment and measurement of event specific behaviors, which by nature suffers from a loss in reliability, even when an experienced clinician's attention is entirely devoted to the process.

All that is typically required for interval scoring is to press down a predefined mouse-, or keyboard-, key when stuttering is perceived. How the stuttering perception works out in terms of the judgment intervals is the task of the computer and, thus, occurs automatically in the background. Because of these features, interval-scoring is likely to become a popular option for many clinicians, especially those who because of treatment related demands, are in need to measure severity frequently.

Possible methodological limitations of interval based stuttering judgments

The simplicity of interval-based judgment measures in case of stuttering, unfortunately, also suggests their potential drawback. That is, its validity in capturing and representing the severity of a stuttering problem may be limited. As all that is measured is a proportion of intervals that are "judged to contain stuttering", this could be all that changes in treatments. The measure does not specifically reflect, for example, how often stuttering behaviors occur, how long they last, and how they affect the overall naturalness of a client's speaking behavior, and levels of tension and effort associated with them. Furthermore, possible variations in speech rate are not reflected by judgments made at the level of scoring intervals. Therefore, if these features are essential for particular treatment programs, a clinician should include additional assessment procedures for reflecting these clinical characteristics. Of course, this in turn would complicate interval scoring to the point where it looses its practical advantages.

There are a number of additional currently unanswered questions about the scaling properties of results obtained with interval based judgment procedures. That is, when sampling periods are relatively brief in duration, a unit of measurement (which is a discrete and fixed interval duration) can represent relatively large steps on the measurement scale. It is not unusual with interval based measures used on speech samples of around 2 minutes, to see that severity values move up or down the measurement scale in steps of 5% or more. The possible interaction of stuttering frequency and duration on stuttering severity is also lost in judgments that are made only relative to entire speaking intervals of a fixed duration. When an interval is marked as stuttered, the result would be the same when it represents one brief instance of stuttering, a multitude of brief stutters during this same measurement epoch, or, in contrast, one long stuttering

block that barely fits in this same scoring interval. How such differences relate to variations in severity such as it is subjectively experienced by the client, and clinician, is presently unknown. One can easily ascertain that the aforementioned behavioral patterns would be perceptually different in these cases, while these behavioral patters would not affect the resulting interval based severity values. One might argue that such misrepresentations may be resolved over time when such judgments are made throughout an entire speech sample. Even so, such robustness has not been empirically studied so far.

In addition to a potential validity loss in representing stuttering severity, interval based procedures are by nature an anemic representation of the stuttering severity such as it is experienced by the stutterer, and by experienced clinicians. Importantly this is so, even though such experiences may not have lead to reliable decisions for treatments. Thus, interval scoring is an attractive, but potentially risky, addition to the clinical armamentarium of stuttering therapists. It's potential use should be carefully considered for the individual case at hand. Further study of the relationship between interval based severity scores and other, reliable and objective estimates of stuttering severity is badly needed but severely hampered by the apparent lack of reliability of even expert observers in making event specific identifications of stuttering dysfluency.

References

Bakker, K., and Gregg, B., (1994). Speech rates based on instrumented, or manually determined, articulation times. *Annual Convention of American Speech and Hearing Association*, New Orleans.

Bloodstein, O., (1995). *A Handbook on Stuttering*. Singular, San Diego.

Bloodstein, O., (1944). Studies in the psychology of stuttering: XIX. The relationship between oral reading rate and severity of stuttering. *Journal of Speech Disorders, 9*, 161-173.

Conture, E.G., & Caruso, A.J., (1987). Assessment and diagnosis of childhood stuttering. In L. Rustin, H. Purser, and D. Rowley (eds.), *'Progress in the treatment of fluency disorders'*, Taylor and Francis, London.

Cordes, A. K., & Ingham, R. J., (1994). The reliability of observational data: II. Issues in the identification and measurement of stuttering events. *Journal of Speech and Hearing Research, 37*, 279-294.

Costello, J. M., & Ingham, R. J., (1984). Assessment strategies for stuttering. In R. J., Curlee, & W. H., Perkins (eds.). *The Nature and Treatment of Stuttering: New Directions*, pp. 303-333. San Diego, CA: College-Hill Press.

Curlee, R. F., (1981). Observer agreement on disfluency and stuttering. *Journal of Speech and Hearing Research, 24*, 595-600.

Ham, R. (1989). What are we measuring? *Journal of Fluency Disorders, 14*, 231-243.

Hegde, M. N., (1995). Measurement and explanation of stuttering: a retrospective appreciation of Gene Brutten's contribution. *Journal of Fluency Disorders, 20*, 205-230.

Johnson, W. & Colley, W. H., (1945). The relationship between frequency and duration of mo-

ments of stuttering. *Journal of Speech Disorders, 10*, 35-38.
Johnson, W., & Associates. (1959). *The onset of stuttering*. Minneapolis: University of Minnesota.
Kazdin, A. E. (1982). *Single-case research designs*. New York: Oxford University Press.
Kully, D., & Boberg, E. (1988). An investigation of inter clinic agreement in the identification of fluent and stuttered syllables. *Journal of Fluency Disorders, 13*, 309-318.
Martin, R. R., & Haroldson, S. K., (1981). Stuttering identification: Standard definition and moment of stuttering. *Journal of Speech and Hearing Research, 24*, 59-63.
Onslow, M., Gardner, K., Bryant, K. M., Stuckings, C. L., & 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.
Perkins, W. H., Kent, R. D., & Curlee, R. F., (1991). A theory of neuropsycholinguistic function in stuttering. *Journal of Speech and Hearing Research, 34*, 734-752.
Riley, G. (1972). A stuttering severity Instrument for children and adults. *Journal of Speech and Hearing Disorders, 37*, 314-320.
Riley, G. D., (1994). *Stuttering severity Instrument for children and adults*.Third edition. Austin, TX: Pro-Ed.
Shaw, C. K., & Shrum, W. F. (1972). The effects of response-contingent reward on the connected speech of children who stutter. *Journal of Speech and Hearing Disorders, 37*, 75-88.
Starkweather, C. W., (1993). Issues in the Efficacy of Treatment for Fluency Disorders. *Journal of Fluency Disorders, 18*, No. 2&3, 151-168.
Young, M. A., (1984). Identification of stuttering and stutterers. In R.F. Curlee and W.H. Perkins, (Eds), *Nature and treatment of stuttering: New Directions*. San Diego: College-Hill Press.
Zebrowski, P. M., (1994). Duration of sound prolongation and sound/syllable repetition in children who stutter: Preliminary observations. *Journal of Speech and Hearing Research, 37*, 254-263.