DISFLUENCY MEASURES: A COMPARISON OF THE ACCURACY OF RANDOM SAMPLING PROCEDURES TO TOTAL COUNT MEASURES

by

LINDA VAN LOENEN

B.S., Kansas State University, 1974

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1975

Approved by:

Major Professor

ACKNOWLEDGMENTS

I wish to express my gratitude to Professor Bruce C. Flanagan for his guidance in preparation of this thesis. A special acknowledgment is also given to the other members of my committee: Professor Norma D. Bunton, Head of the Department of Speech and Professor Thomas M. Longhurst.

My colleague, Janice Parr, deserves special thanks for her assistance in tabulating reliability data in this research study.

A very special thanks is extended to my parents, Mr. and Mrs. Buford Van Loenen for their continuous encouragement and financial support throughout my college experience.

TABLE OF CONTENTS

INTRODUCTION

Psychological rating scale methods, which in the case of stuttering are based on listener rating, and counting moments of stuttering have been predominantly used to measure the degree of stuttering.

Lewis and Sherman (1951) developed one of the first rating scales which consisted of recorded samples of stuttered speech. Each speech sample was represented by a point on an equal appearing interval scale. The Lewis and Sherman scale consisted of eight equal appearing intervals, ranging from 0.50 to 1.49, non-severe, to 7.50 to 8.49, most severe. A study conducted by Berry and Silverman (1971), however, reported that the intervals were "equal appearing" because the observers were instructed to make their ratings in this manner. Therefore, they suggested that it was not safe to assume that the intervals on equal appearing scales were subjectively equal.

A filmed audiovisual scale was developed by Starbuck (1954). Starbuck sequentially arranged samples of stutterers in the act of stuttering to represent the differing degrees of severity. Van Riper (1971) compared the Lewis and Sherman scale with the Starbuck scale and observed that for actual clinical comparisons, both tests were difficult because of the complexity in matching one stutterer against another.

Shumak (1955) compared the stutterers' personal recorded ratings of degree of stuttering with a clinicians' rating of degree of stuttering. The stutterers rated the speech according to a 5-point scale to rate various speaking situations according to four modes of response: avoidance, reaction, stuttering topography and frequency. Ratings of severity of stuttering were also made by clinicians using a 7-point rating scale.

Relatively low coefficients of correlations were obtained, ranging from +.10 to +.54.

Perhaps the most widely used scale for rating severity of stuttering is the Iowa Rating Scale of Stuttering Severity devised by Johnson, Darley and Spriestersbach (1963). It consists of seven scale values which are defined in some detail. Although many speech clinicians use it for diagnosis and for measuring therapeutic progress, Johnson strongly stresses that it should be regarded only as a rough measure of severity.

A comparison of the results obtained when seven different types of scales were used was conducted by Cullinan, Prather and Williams (1963). Twenty-seven recorded samples of speech were rated by a different group of undergraduate students for each of the seven ratings. The resulting data indicated that with only one judge, no one method was reliable enough to use in individual predictions. Results also indicated that none of the seven methods appeared to be any more reliable than the others. Although reliability coefficients of increasing reliabilities can be expected if the mean single ratings from a number of judges is taken, it is difficult under many conditions to obtain such information, due in part to the limited number of clinicians available in a diagnostic and treatment setting.

Riley (1972) and Wertheim (1972) have recently developed rating scales which attempt to eliminate some of the subjectivity of rating scales. Riley designed the Stuttering Severity Instrument (SSI) to yield a single numerical representation of severity within a range of 0 to 45 for use in both clinical and research settings. A point system, scored by the evaluator, measures frequency, duration and physical concomitants. Wertheim proposes a quantitative, multi-dimensional approach which provides information on the situation, stability of the disorder, its qualitative pattern, its severity and the interrelationship between these parameters for a given individual. Although both Riley and Wertheim refer to their measures as objective, subjectivity is required on the part of one or more listeners for both scales.

Young and Prather (1962), who considered the amount of time and labor required to rate speech samples of 200 to 300 words "prohibitive," examined the efficiency of measuring short segments of recorded speech to determine the degree of stuttering. Each speech sample was recorded and divided into successive 20-second segments in order to prepare two segment-test-tapes, the consistently-selected-segment tape and the randomly-selected-segment tape. The second of the segments was clipped from each total sample for the consistently-selected-segment tape. A segment was then selected randomly from the remaining segments of each total sample for the randomly-selected-segment. Forty listeners from a previous study (Young, 1961) rated the total speech samples of 50 stutterers while 14 clinicians rated the 20-second segments on a 9-point scale with little definition of points and with no training in rating being given. After the between-listener variance had been removed, correlations indicating the degree of agreement among raters were 0.81 for the total speech sample, 0.78 for the randomly-selected sample and 0.74 for the consistently selected sample. The authors, due to the relatively low agreement among the raters, suggested that a randomly selected segment of speech be regarded as a reasonable representative average of the total samples only for experimental purposes.

Since distortion and error can result from interpretive characterization of performance when using rating scales, Allyon and Azrin (1968) recommended describing the behavior in specific terms that require a minimum of interpretation. By requiring the observer to attend to and record only in terms of the physical aspect of the response, observer interpretations of the data can be avoided. Several methods of measuring disfluencies have been established to fulfill the need for more descriptive procedures.

Johnson (1961) developed the Iowa Speech Disfluency Test which measures disfluencies according to the following speech behavior classifications: interjections of sounds, syllables, words or phrases; part-word repetitions; word repetitions; phrase repetitions; revisions, incomplete phrases; broken words and prolonged sounds. Samples of the oral reading and speaking of 100 male and 100 female adult speakers, of which 50 in each group were classified by themselves and others as stutterers and 50 as non-stutterers, were measured for speech and disfluency rate. Three speaking tasks, the Job Task, responses to Thematic Apperception Test Cards, and Oral Reading, were performed by each subject. In each case, the disfluencies were identified from a verbatim transcript while listening to a playback of the recording. To insure an accurate identification of disfluencies, the recordings were replayed as often as necessary. Verbal output in each task was defined as the number of words spoken. By calculating the ratio of verbal output to reading or speaking time, as measured by a stopwatch, the speaking and reading rates were computed in terms of words per minute. A disfluency Category Index was made from computation of the number of instances of each type of disfluency per 100 words for each of the three tasks. The formula used to compute this index was Disfluency Category Index=(ND/NW)100, in which ND represents the total number of instances of disfluencies of the designated type in the speech sample of the subject, and NW represents the number of words, or verbal output, for the subject for the sample. By obtaining the sum of each subject's category indexes of a given task, a disfluency index was computed for each task. Johnson reported his data in terms of tables for the following: verbal output on the Job and TAT tasks; measures of time spent on each task; ranges and deciles of distributions of oral reading and speaking rates for the three tasks; measures of disfluency; and coefficients of correlation among the three tasks for measure of each of the disfluency variables of the rate. These findings were then utilized as normative and comparative data for the Iowa Speech Disfluency Test, with respect to rate and disfluency in the speech and oral reading of adult male and female stutterers and nonstutterers.

Young (1961) modified Johnson's disfluency scheme, using five rather than eight categories. These included interjections, part-word repetitions, word-phrase repetitions, prolongations and revisions. A measure of duration, in seconds, was also taken of each tape recorded speech sample. The 50 experimental speakers were required to speak for at least three minutes using Johnson's Job Task to elicit speech. A fluency analysis of the 50 tape recorded samples was then made. This consisted of a 200 word, verbatim transcript and identification and classification of the disfluencies. Each sample could be replayed as often as necessary to ensure sufficient accuracy. A frequency count of the observed occurrences of each of the five types of disfluency, and a measure of speaking time for each of the 50 speakers was computed. Forty-eight experimental listeners, divided into three subgroups, rated the severity of stuttering of each speech sample on the basis of a 9-point equal-appearing intervals scale. A multiple correlation coefficient of .89 and a standard error of the estimate of .99, based on part-word repetitions plus prolongations and time as the predictor of rated severity of stuttering, were obtained. Young's data included a summary of the five measures of disfluency, a frequency count for each type of disfluency, a measure of speaking time, and total frequency of all types of disfluency expressed in terms of mean frequency, range, and standard deviation.

Sanders (1961) investigated the temporal reliability of the Iowa Speech Disfluency Test. This test, containing both reading and speaking tasks, was administered to a group of 40 college-age stutterers and was then repeated 24 hours later. Sanders considered a word disfluent if it involved prolonged sounds, was classified as a broken word; was involved in a sound, syllable or word repetition; or was interrupted by an interjection. He did not count words preceded by interjections or involved in phrase repetitions as disfluent words. The same reading passage was used for both the initial and subsequent administration of the reading task. The speaking situation was also held constant for both days. Results indicated the coefficient of correlation to be .96 for total disfluent words. Sanders stated that these results were for total number of disfluencies only. It does not indicate the extent of agreement for the individual disfluency categories or for the occurrences of particular disfluencies. A test-retest Pearson product-moment correlation coefficient was computed for both tasks resulting in scores of .91 and .94 total disfluencies on the speaking and reading tasks respectively. These results, indicating rather high temporal reliability, suggest the feasibility of using such a method to measure disfluencies.

Each of these studies (Johnson 1961, Young 1961 and Sanders 1961) measured rate of utterance and classified disfluencies into four to eight classifications. Hammond (1973) stated that these methods were time consuming and tedious, requiring from two to four hours to transcribe and analyze. Johnson (1961) stated that much time was required, often several hours, to obtain an accurate transcription and identification of the disfluencies for each individual speech sample. This limitation, analysis and categorization of disfluencies for 300 word speaking samples and computation of utterance, would indicate a need for a more efficient procedure which affords comparable accuracy.

Realizing the need for a valid and efficient method for disfluency analysis, Hammond (1973) studied the relationship between the Time Interval Disfluency Test (TIDT) and Sander's Disfluent Word Index. Each four minute TIDT speaking task, reading, monologue and conversation, was recorded simultaneously with 10-second beeps which sub-divided the sample into 10-second intervals. Listeners marked a score sheet to indicate if the client was speaking, if he was disfluent or if the interviewer was speaking. A total count of disfluencies was not made. The number of intervals in which a disfluency occurred was then divided by the total number of intervals in which the subject was speaking to obtain a percent disfluent score. The total number of 10-second intervals in which the client was speaking was then divided by the total number of intervals available to obtain an estimate expressed as a percent of speaking time.

The Sander's Disfluent Word Index, consisting of a reading and Job Task, yielded a percentage of disfluent words by dividing the total number of disfluent words identified by the total number of words. The rate of speaking time was computed by counting the total number of seconds for each speaking task. Nine male stutterers performed both the two Sander's tasks and the three TIDT speaking tasks. Coefficients of correlations for the mean of the judges were as follows: .82 when comparing the TIDT reading task to Sander's reading task; .82 when comparing the TIDT reading task to Sander's Job Task; .29 when comparing the TIDT monologue task to Sander's reading task; .49 when comparing the TIDT monologue task to Sander's Job Task; .39 when comparing the TIDT conversation task to Sander's reading task; .63 when comparing the TIDT conversation task to Sander's Job Task. These correlations may indicate that the TIDT is measuring behaviors in addition to those measured by Sander's Disfluent Word Index. These correlations indicate that each task of the TIDT may be independent measures. Hammond (1973) stated that "each is a separate entity, each with a place in the diagnostic evaluation." In terms of administration, the TIDT scoring procedure required a score sheet and 15 minutes per subject while Sander's procedure required a verbatim transcript and two to four hours per subject.

Buetzer (1973) conducted a study to determine the reliability of the Time Interval Disfluency Test on an inter-observer agreement basis. Training sessions were utilized immediately prior to each analysis session, to facilitate accurate listener analysis of the speech samples. During the training sessions, the eleven listeners were required to reach 80 percent agreement with the master score sheet developed by the author and two assistants. Scorer reliability was indicated by percentage of agreement of each observer to all other observers in terms of whether a 10-second interval was marked or unmarked and coefficients of correlation of the total disfluency percentage computed for each of the speaking tasks from each subject by each of the eleven observers. Results of the median coefficient of correlation of agreement among subjects by tasks were: .88 for the reading task; .79 for the conversation task; and .79 for the monologue task. These coefficients, significant at the .01 level of confidence, indicated significant reliability of the TIDT on an inter-observer agreement basis.

Although the Time Interval Disfluency Test has been shown by Hammond and Buetzer to be a rapid, efficient measurement of degree of stuttering, it is insensitive to certain clinical phenomenon. From the current author's experiences and those of other clinicians it is apparent that a speaker could receive a high percentage for fluency even though he produced a number of disfluencies in a single interval. For example, if the client produced 40 disfluencies in the first two 10-second intervals, only one disfluency would be recorded in each segment. Viewing this problem from the opposite extreme, a client who produced only a single disfluency in each of the first two 10-second segments would receive the same percentage for fluency as in the previous example. Few listeners, however, would attach similar ratings of degree of stuttering to such examples. With this system, only an estimate of degree of disfluency is obtained since an accurate count of number of disfluencies, total number of words spoken by the clinician and total words spoken by the client are not computed. A further criticism of this procedure is based on the amount of speaking time computed for the client and for the clinician. Using this system, the client is required to produce only a single-word utterance to be scored as fluent, or disfluent, during a given 10-second interval. Theoretically, a client who produced one fluent word during an interval would receive the same score as a client who produced 20 to 30 fluent words during a similar interval. It is obvious, however, that the greater the amount of time the client is engaged in talking, the greater are his opportunities to be disfluent and the more relevant are his scores. Using the TIDT, however, both types of intervals would receive identical scores. An inaccurate, high percentage of client speaking time and of fluency could be scored from limited client response.

The focus of this study will be to provide a method to measure frequency of stuttering and words spoken which is both rapid and sensitive to the aforementioned clinical phenomenon. The purpose of this study was to determine the number of 15-second segments which must be analyzed to obtain reliable measures of disfluency and fluency word counts observed during conversational speech. The conversational speech task was selected since it may be considered most representative of the stutterer's communicative experiences. In support of this posture, Ayllon and Azrin (1968) indicated that only those behaviors that will continue to be reinforced after training be selected for training in a clinical setting. A conversational speaking task, as opposed to the monologue or reading tasks, appears to have a greater probability of being maintained in social settings.

METHOD

Subjects

Twelve males and two females, all of whom were considered by themselves and by two speech pathologists to be stutterers, served as subjects. These disfluent subjects, whose ages were from 12-42 years, were obtained from the Kansas State University Speech Clinic, the University of Kansas Speech Clinic and the adjacent community. All of the subjects had at one time received or were presently receiving treatment for stuttering.

Recording Procedure

Each subject was seated beside the interviewer at a table in a quiet room with a microphone positioned in front of him. Speech samples were recorded using a Wollensak tape recorder, Model number T 1500. The recording equipment was located either in a room adjacent to the experimental room or on the floor beneath the table. A brief interview was conducted to accustom the subject to the experimental situation and to adjust the recording level of the tape recorder. The experimenter and two speech pathology graduate students served as interviewers.

The conversational speech sample was obtained from questions asked by the interviewer to elicit conversational speech. These questions asked for information such as what types of leisure activities the subject engaged in, what field of study he was in, what types of hobbies he had and what places he had or would like to travel to. The interviewer responded appropriately to the subject's conversation, as would occur in a typical conversation. See Appendix A for the verbal prompts provided by the interviewer. These prompts were adopted from those employed by Hammond (1973).

Disfluency Definition

To accommodate a conversational speech sample, a modification of Johnson's (1961) definition of disfluency was adopted. Included from this definition were the following categories: an interjection of sounds, syllables, words, and phrases; part-word repetitions; word repetitions; phrase repetitions; broken words and prolonged sounds. Excessive pauses were also counted as disfluencies (Love and Jeffress, 1971). Revisions and incomplete phrases were not counted as disfluent words because of the judged difficulty in reliably counting such phenomenon in conversational speech.

Data Analysis

A verbatim transcript of the total sample of both interviewer and stutterer responses was written to enable the experimenter to segment the speech sample into 15-second segments and to obtain a word count per sample. Measures were obtained for the number of words spoken by the interviewer, the number of disfluent words spoken by the stutterer and the number of fluent words spoken by the stutterer for each 15-second segment and for the total sample.

To obtain the word count, a modification of Siegel's (1963) criteria for counting words was employed. Included from this criteria were the following categories: expressions of affirmation, of negation, or of exclamation counted as one word; hyphenated words and compound nouns which seemed to function as single words were counted as one word each; and exclamations which tend to occur as a unit were counted as one word. Contractions and combinations such as "gonna" or "hadda" were counted as one word rather than as two words because of the judged difficulty of counting such expressions in conversational speech. Other words repeated singly or in a phrase were counted only once.

The total speech sample then consisted of eight minutes segmented into thirty-two 15-second intervals. Flanagan (1973) projected that a seven to eight minute speech sample would yield test-retest reliability of 90 percent agreement with respect to the frequency of pauses and vocal utterances. The 15-second segments were indicated on the verbatim transcript by a slash mark (/) which corresponded with 15 seconds on the tape recording as measured by a stop watch. All segments in a sample were then sequentially numbered from one to thirty-two.

Groups of varying numbers of randomly selected segments were compiled next to determine the efficiency of measuring varying numbers of short segments of speech. More specifically, the entire speech sample and the following numbers of randomly selected segments were examined: sixteen 15-second segments; eight 15-second segments; four 15-second segments; two 15-second segments; and one 15-second segment.

The random selection process was as follows: thirty-two 3" x 5" note cards were numbered to correspond to the appropriate 15-second segment. For example, card number one corresponded to the first 15-second segment, card number 10 corresponded to the tenth 15-second segment and card number thirty-two, the last card, corresponded to the last 15-second segment. These cards were then shuffled three times, followed by cutting the deck, for each subject by condition. The experimenter then counted off the appropriate number of cards from the top of the deck for each experimental measurement to indicate which 15-second segments were to be examined.

For each 15-second segment, a count was then recorded for the number of fluent words spoken by the stutterer, the number of disfluent words spoken by the stutterer and the number of words spoken by the interviewer. These results were then tabulated for the randomly selected speech samples. A Pearson-product-moment correlation was obtained for each word sample size to assess the reliability of measuring the indicated numbers of randomly selected 15-second segments.

A t-test analysis was also computed to determine the significance of the difference between the means for frequency of stuttering reported for Sander's (1961) study containing forty adult stutterers and the present research study.

Scorer Reliability

All segments were examined by the experimenter. Scorer reliability between the experimenter and another listener, both graduate students in speech pathology, was measured by Pearson's product-moment correlational procedures. To accomplish this for each subject's speech sample, four randomly selected 15-second segments were examined to determine whether the number of disfluent words spoken by the stutterer, number of fluent words spoken by the stutterer and the number of words spoken by the interviewer were similarly recorded. The observer, however, did not obtain the word counts for the word categories from a verbatim transcript. Instead, she was instructed to record the word counts by making hash marks on a sheet of paper for each observed word of a word category. The observer was allowed to replay each 15-second segment as often as necessary to allow her to be satisfied with the accuracy of her counts. A total word count for each word category was then derived for each subject by totaling the number of words recorded by the observer during the four 15-second segments.

Results indicated the inter-observer coefficients of correlation to be .9718 for the number of fluent words spoken by the stutterers, .9028 for the number of disfluent words spoken by the stutterers, and .9933 for the number of words spoken by the interviewers. These coefficients, significant at the .001 level of confidence, indicate significant scorer reliability.

RESULTS AND DISCUSSION

The results of this research are presented in reference to: 1) the t-test analysis for significance of the difference between means for frequency of stuttering for Sander's (1961) study and this research study and 2) the coefficients of correlation obtained when comparing word counts for the total eight minute speech sample to word counts for several randomly selected 15-second segments. Methods are then discussed which would enable a clinician to directly apply this measurement procedure to the clinical environment.

Sander (1961) reported a study to test-retest reliability of the procedure which he developed for the measurement of disfluent words. His definition of disfluent words closely approximated the definition of disfluent words employed in the present study. Sander's sample of stutterers was comprised of 40 adult stutterers, 34 males and 6 females. The present study included 14 stutterers, 12 males and 2 females. A t-test, for unrelated groups, was computed comparing the mean disfluency rate for these two groups. The results of this analysis resulted in a t equal to 0.51 which, utilizing 52 degrees of freedom, suggested the observed mean difference could be obtained by chance 68 times out of 100. These results indicate the mean rate of disfluencies for the stutterers in the present study are not, statistically, significantly different from the mean rate of disfluency for the group of stutterers studied by Sander. The implication of this result is that this research is probably studying a population similar to the one utilized by Sander with respect to rate of disfluency.

Recall that for each 15-second segment, a count was recorded for the number of fluent words spoken by the stutter, the number of disfluent words spoken by the stutter and the number of words spoken by the interviewer during an eight minute conversational speech sample. The word counts were then tabulated for the entire eight minute speech sample (i.e., thirty-two 15-second segments) and for the following numbers of randomly selected segments: sixteen 15-second segments, or a total of 4 minutes of sampling time; eight 15-second segments, or a total of 2 minutes of sampling time; four 15-second segments, or a total of one minute of sampling time; two 15-second segments, or a total of 30 seconds of sampling time; and one 15-second segment. The results obtained for the entire eight minute or thirty-two segments were then compared to the results obtained for each random sample to determine the minimum number of 15-second intervals which must be sampled to maintain a high level of coefficient of correlation with the total speech sample. Table 1 presents the obtained coefficients of correlation.

For the number of disfluent words spoken by the stutterers, the following coefficients were obtained: .93 for sixteen 15-second segments; .96 for eight 15-second segments; .90 for four 15-second segments, .60 for two 15-second segments; and .80 for one 15-second segment.

For the number of fluent words spoken by the stutterers, the following coefficients were obtained: .93 for sixteen 15-second segments; .95 for eight 15-second segments; .90 for four 15-second segments; .76 for two 15-second segments; and .83 for one 15-second segment.

For the number of words spoken by the interviewer, the following coefficients were obtained: .90 for sixteen 15-second segments; .90 for eight 15-second segments; .64 for four 15-second segments; .36 for two 15-second segments; and .70 for one 15-second segment.

These correlations appear to indicate that for all three variables under study, eight randomly selected 15-second segments, or one-fourth of the

TABLE I. Product-Moment Correlations of Word Counts for an Eight Minute Conversational Speech Sample and a Number of Randomly Selected Fifteen Second Samples for Stutterers' Disfluent Words, Stutterers' Fluent Words and Interviewers' Words

Category of	Correlation for Number of 15-Second Segments Sampled				
Word Count	16 Segments	8 Segments	4 Segments	2 Segments	1 Segment
Stutterers' Disfluent Words	.93	.96	.90	.60	.80
Stutterers' Fluent Words	.93	.95	.90	.76	.83
Interviewers' Words	.90	.90	.64	.36	.70

entire conversational speech sample, are the minimum that a clinician can sample to maintain a .90 coefficient of correlation. It is generally agreed that coefficients of correlation of .90 or higher are acceptable for predicting measured results for individual subjects. It could be argued that for the stutterers' disfluent and fluent words a sample size of four 15-second segments, or one minute, would be sufficient for the prediction of an individual's results during an eight minute speech sample. Such an argument could be accepted from a purely logical point of view, given the assertion stated above that a minimum correlation coefficient of .90 is sufficient for prediction of individual results. However, it should be pointed out that a correlational coefficient of .90 is the minimum acceptable value for individual prediction. It is the author's posture that the additional predictive power suggested by the two minute sample (eight 15-second segments) contributes substantially to the reliability of the predictive values for stutterers' disfluent words (r=.96) and stutterers' fluent words (r=.95). Additionally, from a practical stand point, the utilization of the findings of these results in a clinical setting may be confusing. The clinician may misunderstand instructions which state that it is necessary to randomly select four 15-second segments of conversational speech for stutterers' disfluent and fluent word counts and to select eight 15-second segments for the interviewer's word count. Further, clinicians may "tend to be more interested" in the client's behavior rather than in obtaining measures of the interactive communicative context of conversational speech.

For the above reasons, it was concluded that a minimum of eight randomly selected segments of an eight minute conversational sample are adequate to predict the word counts for stutterer's disfluent and fluent words and for the interviewer's words. This conclusion is limited to an eight minute conversational speech sample obtained as described in the previous chapter and, further, to specifically exclude generalization of these findings to oral reading or to speech samples collected according to other procedures.

For the three categories of word counts, the coefficients of correlation generally attenuated as the number of 15-second segments was reduced, except for the one 15-second sample in which the coefficients of correlation appeared to increase for the three counts. The author is unable to account for this unique result except to state that such an increase in these coefficients could occur by chance one time out of eight.

A significant aspect of this research study is that it readily lends itself to a practical and efficient method for evaluating progress in stuttering treatment. This measurement procedure allows the clinician to count three measures related to treatment progress in a relatively efficient manner.

As a clinical tool, this procedure requires a tape recorder with a counting wheel, a stopwatch, a pencil, a sheet of paper and a deck of thirty-two cards each of which indicates a 15-second segment (i.e., 1 second to 15 seconds, 16 seconds to 30 seconds...5 minutes 1 second to 5 minutes 15 seconds...7 minutes 46 seconds to 8 minutes). The clinician must randomly select eight of these cards to determine which 15-second segments to analyze. This can be accomplished by shuffling the thirty-two cards and then selecting the first eight cards. For convenience, these cards can then be ordered by real time (i.e., the lowest time card first).

To obtain a word count for the number of disfluent words spoken by the client, the following definition may be employed: 1) an interjection of sounds, syllables, words or phrases; 2) part-word repetitions; 3) word repetitions; 4) phrase repetitions; 5) broken words; 6) prolonged sounds; or 7) excessive pauses. Extraneous sounds such as "uh," "um," and "er," extraneous words such as "well" and "oh," and extraneous phrases such as "you know" are considered as interjections. A single count of disfluency is recorded regardless of the number of units of repetitions of a single interjection. For example, "uh" and "uh uh uh" are each counted as one instance of disfluency. However, if two different interjections are observed, each is counted as a disfluent word. For example, "uh well" and "uh uh well" are each counted as two instances of disfluencies.

Single counts of disfluency are recorded regardless of the number of repetitions of parts of words, of words or of phrases. For example, "ba-baby," "I-I-I-," and "I was I was I was going" each count as a single instance of disfluency. A word repeated for emphasis, as in "really, really fast," is not counted as a disfluent word. A part-word repetition or an interjection does not nullify a word repetition, as in "baby uh baby" or "buh-baby baby." A disfluency is counted for both the interjection and the repetition.

Words which are not pronounced completely or which are broken in a way that definitely interferes with the smooth flow of speech are considered a broken word. For example, "want to p- (pause) lay" would include a single count of disfluency.

Single counts are also recorded for excessive pauses perceived either during a phrase or at the completion of a phrase. A pause of one-tenth to one-fifth of a second or longer during a phrase is considered a disfluency. Longer pauses at the end of a phrase must be observed to be counted as a disfluency.

To obtain a word count for the number of fluent words spoken by the client and for the number of words spoken by the clinician, the following criteria are recommended: 1) expressions of affirmation ("uh-huh," "yeh"), of negation ("huh-uh," "nah"), or of exclamation ("wow") count as one word; 2) hyphenated words ("step-mother") and compound nouns ("baseball") which function as a single word counts as one word each; 3) exclamations which tend to occur as a unit "oh boy") are counted as one word; and 4) contractions ("isn't") and combinations of words ("gonna," "hadda,") count as one word. Other words repeated singly or in a phrase are counted only once. The clinician is instructed to use her own clinical judgment to consistently count disfluencies or words which were observed during the speech sample but are not contained in either definition.

After the clinician has ordered the cards by real time, she starts both the stopwatch and the tape recorder. She listens to the tape of the conversation until the segment indicated by the first card is reached. The number indicated on the counting wheel which corresponds to this segment on the tape is then noted on the paper. The clinician might also note the words spoken by either the client or the clinician at the beginning of this interval. The clinician then begins recording the number of disfluent words spoken by the client for this 15-second interval. After the 15 seconds have elapsed, the clinician stops both the tape recorder and the stop watch and again notes the number indicated on the counting wheel of the tape recorder. Again, it would benefit the clinician to note the words produced by either the client or the clinician at the end of the 15-second interval. The clinician then re-winds the tape to the number noted on the counting wheel that corresponds to the beginning of the 15-second segment. She then replays that portion of the tape two additional times, once to obtain a count for the number of fluent words spoken by the client and once to obtain a count for the number of words spoken by the clinician. If necessary the tape can be replayed to insure sufficient accuracy for each type of category.

All three word counts can be recorded on paper by making a hash mark for each observed word by measurement category. The hash marks are then totaled and recorded for each of the word categories in a 15-second segment. Once the clinician has determined a count for each word category, the stopwatch and tape are again started and the clinician listens to the tape until the next 15-second interval is indicated. The clinician again notes the number indicated on the counting wheel and the first words spoken for the interval. A count is then made for the number of disfluent words spoken by the client during this second 15-second segment. The clinician stops both the tape recorder and the stop watch at the completion of the segment and notes the number indicated on the counting wheel and the words spoken at the end of the segment. The interval is then replayed and word counts are compiled for the number of fluent words spoken by the client and the number of words spoken by the clinician. This procedure is to be repeated until all eight 15-second segments have been properly recorded. At the completion of the eighth 15-second segment, the clinician rewinds the tape and tabulates the word counts obtained from all eight 15-second segments for each word category. These three total word counts for each word category are then treated as representative of the results obtained from measurement of the eight minute conversational speech sample.

The clinician can consider these three word counts, number of fluent words spoken by the client, number of disfluent words spoken by the client, and number of words spoken by the clinician, to initially constitute a baseline of the client's fluency and speech production during conversational speech. Once the clinician has determined the baseline for these word counts, an intervention program can be instituted. The clinician then utilizes this measurement procedure as a probe to determine the effect of the speech therapy program on the client's typical speech fluency. Ideally, these probes should be taken in a neutral environment, or someplace other than the therapy room, and with a neutral listener, or someone who hasn't been included in the training program. Obviously this optimum probe will be difficult to obtain in a clinical situation and compromises might be made. Perhaps the clinician could consistently tape the first eight minutes of the first therapy session of the week. This would enable the clinician to measure the client's fluency after a weekend without therapy and before the client is again required to specifically utilize the therapy training procedures. Other suggestions include obtaining the speech sample while walking with the client in the hall or while the client is visiting with someone in the waiting room prior to therapy (using miniaturized portable recording equipment).

Regardless of the situation from which the speech sample was taken, the speech sample should contain conversation which requires the client to respond to open-ended questions, or questions which cannot be answered by the client with a simple "yes" or "no" answer. Examples of such questions are "What types of leisure activities do you engage in?", "Where would you like to go if you had the opportunity? Why?", or "What would you like to change about your schooling so far?"

To avoid biasing the results and to obtain a representative sample of the client's typical speech, the client should not be sensitive to the fact that a probe is being obtained. This can be achieved by continuous taping of the entire probe and treatment session. Other legally acceptable methods of obtaining accountability data may also be used.

After word counts for each word category have been compiled for the entire two minute speech sample, several alternative methods of analysis are available to the clinician. The clinician may utilize the results to measure the client's vocal output by 1) computing the number of words spoken per minute for each word category; 2) computing a percentage of fluency or of disfluency; 3) graphically displaying the results for each word category on linear chart paper; or 4) graphically displaying the results for each word category on six-cycle chart paper. Although the author prefers the fourth alternative, graphically recording the results on six-cycle chart paper, a brief discussion of each alternative procedure will be presented with greater emphasis on this fourth alternative.

By computing the number of words spoken per minute for each word category, the clinician can obtain a measurement of rate of fluency (Client's Fluent Word Count/2), and the rate of disfluency (Client's Disfluent Word Count/2). The clinician can then compare the rates obtained for each word category to rates obtained from previous probes to assess improvement in therapy. The clinician can also compare the results for words spoken per minute by the clinician to the total words spoken per minute by the client to indicate the effect of her own verbalization during the probe.

The second alternative allows the clinician to assess the client's speech by computing a percentage of fluency (Client's Fluent Word Count/Client's Total Word Count) or by computing a percentage of disfluency (Client's Disfluent Word Count/Client's Total Word Count). Goals for therapy will determine which percentage to compute. Again, results obtained from each probe can be compared to indicate improvement during treatment.

The clinician might choose to graphically represent the word counts obtained for each word category on a linear chart, the third alternative. By graphing these results, the clinician can visually compare the client's performance during each probe. Percentages of fluency or disfluency and rate of words spoken per minute can also be graphed on a linear chart. However, a separate chart would be required for each type of analysis.

The fourth alternative, graphically displaying the results for each word category on six-cycle chart paper, allows the client to visually measure on a single chart the rate of utterance for each word category, the accuracy ratio and the percentage of fluency. Six-cycle chart paper differs from the traditional graph paper by the inclusion of time as a variable. Although relatively little information is available on the utilization of six-cycle chart paper in speech pathology, it has received considerable support from proponents of precision instruction.

In order to plot this data and to understand the information available from the chart, a basic knowledge of charting principles is required and will be presented in the following discussion. For a more detailed explanation of charting principles, the clinician should consult the Handbook of the Standard Behavior Chart by Penneypacker, Koenig and Lindsley (1972).

Basically the six-cycle chart measures frequency in terms of counts per minute (C/m). The lowest frequency that can be observed is the reciprocal of the number of recording minutes. This value, called the sample floor, is calculated as 1/No. of Minutes Spent Recording. For this measurement procedure, the sample floor is 1/2 minutes or .5 on the chart.

However, if the clinician measures only two minutes, or one-fourth of the total eight minute speech sample, a record floor should also be included on the chart. The record floor indicates the total number of possible minutes from which the sample was selected. It is defined as 1/No. of Minutes the Behavior Can Occur. For this measurement procedure, the record floor is 1/8 minutes or .125 on the chart. The sample floor then tells how many minutes the behavior was actually sampled while the record floor tells the number of minutes the behavior could have occurred and therefore could have been recorded. Those minutes in which the behavior was occurring but was not being recorded is referred to as ignored time.

Sample floor, record floor and ignored time are depicted on a chart using the following symbols: a series of horizontal dashed lines connecting Tuesday and Thursday represent the sample floor; a solid horizontal line represents the record floor and a series of diagonal lines between the sample floor and the record floor represent the ignored time.

To rapidly and accurately plot the frequencies for the three word counts obtained from this research procedure, a clinician should employ the use of a "frequency finder." By laying the frequency finder alongside the frequency scale of a chart, a clinician can rapidly plot both frequencies and floors.

To plot the sample floor, the clinician makes a dot on the frequency finder corresponding to two, or the number of minutes spent recording. This dot is then placed on the 1 C/m line of the chart. The arrow on the frequency finder then indicates the sample floor and the clinician marks a series of dashed lines along this frequency of the chart.

To plot the record floor, the clinician makes a dot on the frequency finder corresponding to eight, or the number of minutes the behavior could have been sampled and recorded. Again, the dot is placed on the 1 C/m line of the chart. The record floor is indicated by the arrow on the frequency finder and the clinician draws a solid line corresponding to this point.

To plot the frequency for the three measured word counts, the following symbols should be utilized: a dot represents the number of fluent words spoken by the client; an X represents the number of disfluent words spoken by the client; and a / represents the number of words spoken by the clinician.

To plot the correct frequency for each measured word count, the clinician must first locate the sample floor. Then, with the 1-line of the frequency finder on the sample floor, the clinician finds the line on the frequency finder which corresponds to the number of words computed for each measurement. The clinician then places the correct symbol for the measurement on the chart. For example, if the clinician observed ten disfluencies during the two minute sample, the arrow on the frequency finder would be aligned with .5, the sample floor. The clinician would then locate the 1-line on the frequency finder and place an X on the chart opposite this line. If the clinician counted 200 fluent words spoken by the client during the two minute sample, the arrow on the frequency finder would again be aligned with .5, the sample floor, on the chart. The clinician would then locate the 200-line on the frequency finder and place a dot on the chart opposite this line. If the clinician counted 150 words spoken by the clinician during the two minute sample, the arrow would again be aligned with the sample floor, .5 on the chart. This time the clinician would locate the 150-line on the frequency finder and place a / on the chart opposite this line. If no instances of a word count occurred during the period of time recorded, the clinician would place the appropriate symbol on the proper day line just beneath the sample floor.

Since the frequency scale of the six-cycle chart paper is an equal ratio scale, all distances along it represent ratios or proportions. Therefore, the clinician can convert the distance between the number of disfluent words spoken by the client to the number of fluent words spoken by the client into a proportion or percentage. To determine proportions between these two word counts, the clinician must align the arrow, the 1-line, of the frequency finder with the X which indicates the number of disfluent words. By reading the value on the frequency finder which corresponds with the dot representing the number of fluent words spoken by the client, the clinician can compute accuracy pairs. The resulting accuracy ratio indicates the proportion of disfluent to fluent words.

To determine the percentage of fluency, the clinician aligns the arrow on the 1-line of the frequency finder with the dot which indicates the number of fluent words spoken by the client. The clinician then reads the value on the frequency finder which corresponds to the X indicating the number of disfluent words spoken by the client and multiplies it by 100. The resulting number then indicates the percentage of fluency.

By charting the data on six-cycle chart paper, all ratios and percentages computed in reference to the two minute sampled period remain constant for the entire eight minute speech sample.

A clinician, however, is interested not only in analysis of daily data results. The client's growth or improvement during therapy is another factor which is of importance to the clinician. When measuring fluency, the client can improve by increasing the number of fluent words spoken, an increase in speed, or by decreasing the number of disfluent words spoken, an increase in accuracy. Various combinations of changes in speed and accuracy may occur and can be observed from the six-cycle chart. By examining the number of words spoken by the clinician, the clinician can also determine the effect of her own verbal output on the client's speech during the conversational speech sample.

Although this fourth method of analysis is preferred by the author, each alternative would provide considerable therapeutic information to the clinician concerning progress in stuttering therapy. Other possible methods of analysis could also be explored by the clinician.

REFERENCES

Ayllon, T. and Azrin, N. H. The token economy: a motivational system for therapy and rehabilitation. New York: Appleton-Century-Crofts, Inc., 1968.

Berry, R. and Silverman, F. Equality of intervals on the Lewis-Sherman Scale of Stuttering Severity. Journal of Speech and Hearing Disorders, 1972, 15, 185-188.

Buetzer, C. B. Reliability of the Time-Interval Disfluency Test on an inter-observer agreement basis. Unpublished Master's thesis, Kansas State University, 1973.

Cullinan, W. L., Prather, E. M., and Williams, D. E. Comparison of procedures for scaling severity of stuttering. Journal of Speech and Hearing Research, 1963, 1, 187-193.

Flanagan, P. Durational characteristics of oral reading: effects of sample size on test-retest reliability. Unpublished Master's thesis, Kansas State University, 1973.

Hammond, C. A study of the interrelationships between the Time-Interval Disfluency Test and Sander's Disfluent Word Index. Unpublished Master's thesis, Kansas State University, 1973.

Johnson, W. Measurements of oral reading and speaking rate and disfluency of college age male and female stutterers and nonstutterers. Journal of Speech and Hearing Disorders, 1961, Monograph Supplement 7, 1-20.

Johnson, W., Darley, F. L. and Spriesterbach, D. C. Diagnostic methods in speech pathology. New York: Harper & Row, 1963.

Lewis, D. and Sherman, D. Measuring the severity of stuttering. Journal of Speech and Hearing Disorders, 1951, 16, 320-326.

Love, L. R. and Jeffress, L. A. Identification of brief pauses in the fluent speech of stutterers and nonstutterers. Journal of Speech and Hearing Research, 1971, 14, 229-240.

Pennypacker, H. S., Koenig, C. H., and Lindsley, O. R. Handbook of the Standard Behavior Chart. Kansas City, Kansas: Precision Media, 1972.

Sander, E. K. Reliability of Iowa Speech Disfluency Test. Journal of Speech and Hearing Disorders, 1961, Monograph Supplement 7, 21-30.

Siegel, G. M. Prototypes of criteria for counting words. Journal of Speech and Hearing Disorders, 1963, Monograph 10, 103.

Shumak, C. A speech situation rating sheet for stutterers. In W. Johnson (ed.) Stuttering in Children and Adults. Minneapolis: University of Minneapolis Press, 1955.

Starbuck, H. B. Determination of severity of stuttering and construction of an audio-visual scale. Doctoral dissertation, Purdue University, 1954.

Van Riper, C. The nature of stuttering. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1971.

Wertheim, E. A new approach to classification and measurement of stuttering. Journal of Speech and Hearing Disorders, 1972, 37, 242-251.

Young, M. A. Predicting ratings of severity of stuttering. Journal of Speech and Hearing Disorders, 1961, Monograph Supplement 7, 31-54.

Young, M. A. and Prather, E. M. Measuring severity of stuttering using short segments of speech. Journal of Speech and Hearing Research, 1962, 5, 256-262.

APPENDIX A

QUESTIONS USED TO ELICIT CONVERSATION

1. What types of leisure activities do you engage in?
2. Have you ever participated in any organizations?
3. How active are you in political affairs?
4. How would you describe your personality characteristics?
5. Are you living on campus (on the army base)? Do you like it?
6. How relevant do you feel your education was for you?
7. Do you feel that coming to KSU (KU or Fort Riley) has affected your attitudes in any way?
8. What would you have liked to change about your schooling so far?
9. What field of study are you in?
10. After you obtain your degree are you planning on graduate study?
11. Do you have any brothers and sisters? Can you tell me about them?
12. How would you describe your relationship with your parents?
13. What places have you traveled to?
14. Where would you like to go if you had the opportunity? Why?
15. What types of things really irritate you about people or situations?
16. What particular traits do you admire in people?
17. If there was one thing you could change about yourself what would it be?
18. What do you feel your best qualities are?
19. What person have you most admired? Why?

DISFLUENCY MEASURES: A COMPARISON OF THE ACCURACY OF RANDOM SAMPLING PROCEDURES TO TOTAL COUNT MEASURES

by

LINDA VAN LOENEN

B.S., Kansas State University, 1974

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1975

The purpose of this research was to determine the minimum number of 15-second segments which must be analyzed to maintain high coefficients of correlation with the total eight minute conversational speech sample for word counts of stutterers' fluent and disfluent words and of interviewer's words.

Tape recorded conversational speech samples were obtained from twelve male and two female stutterers. The eight minute speech sample was then divided into thirty-two 15-second segments. The entire speech sample and the following numbers of randomly selected segments were examined: eight 15-second segments; four 15-second segments; two 15-second segments; and one 15-second segment.

For each 15-second segment, a count was recorded for the number of fluent words spoken by the stutterer, the number of disfluent words spoken by the stutterer, and the number of words spoken by the interviewer. These results were tabulated for the entire speech sample and for the randomly selected samples.

Scorer reliability between the experimenter and another listener was accomplished by examining four randomly selected 15-second segments from each speech sample. A total word count for the four 15-second segments was then derived for each word category for each subject. All three measured word counts maintained a .90 coefficient of correlation, indicating a significant inter-observer reliability.

Data analysis consisted of Pearson's product-moment correlation to compare word counts obtained from the entire conversational speech sample to results obtained from each group of randomly selected 15-second segments. A t-test analysis was also employed to determine the significance of the difference between the means for frequency of stuttering reported for

Sander's (1961) study containing forty adult stutterers and for the present research study.

Results of the coefficients of correlation appear to indicate that for all three variables under study, eight randomly selected 15-second segments, or one-fourth of the total conversational speech sample, are the minimum which a clinician can sample to maintain an acceptable level of prediction ($r=.90+$).

Results of the t-test indicated that the mean rate of disfluencies for the stutterers in the present study are not statistically, significantly different from the mean rate of disfluency for Sander's group of stutterers.

The measurement procedures employed allows the clinician to count three measures related to treatment progress of stutterers in a relatively rapid and efficient manner. The discussion focused on how these findings may be utilized by the clinician in the treatment evaluation for stutterers. Alternative methods suggested for analysis of resulting word counts include: 1) computing the number of words spoken per minute for each word category; 2) computing a percentage of fluency or of disfluency; 3) graphically displaying the results for each word category on linear chart paper; and 4) graphically displaying the results on six-cycle chart paper.