

# THE RELATIONSHIPS BETWEEN MEASUREMENTS OF STUTTERING BEHAVIOUR

MYRTLE L. ARON, PH.D.

*Department of Logopedics, University of the Witwatersrand, Johannesburg*

In our attempt to ascertain the symptomatology of stuttering and the change and progress that might occur in particular situations and in therapy, evaluative procedures become essential. Measurements of the dimensions of stuttering behaviour is part of the evaluative procedure. Not all the dimensions have been established—there are many facets which remain elusive in terms of objective assessment. Stuttering varies a great deal under certain known conditions but it can also vary for no apparent reason. The cyclic phenomenon in stuttering has been discussed by Quarrington,<sup>11</sup> and others, but it has not been explained satisfactorily. An objective study of variability in stuttering is therefore difficult to make as any change that may occur may not necessarily be associated with manipulated conditions or the process of therapy. Individual personality differences add to the problem of objective measurement in stuttering. However, these factors need not preclude our use of measurements and some attempt must be made to quantify data. This is not to say that we must belittle qualitative judgements. These finally are the most important. In many respects those measurements we can make at the present time reflect qualitative assessments. The need to attempt measurement, although we may not have all the answers, is supported by Thurstone<sup>18</sup> who states that:

... it is better to formulate the law of comparative judgements in terms of the discriminial error, which is a psychological concept, than to wait until we shall understand physiologically . . .

While conducting a study to ascertain the effects of a combination of tranquillizing and sedatory agents on the symptoms of stuttering, an array of measurements were used to assess stuttering speech. It is not the intention here to report on the effects of the drug as this has been considered elsewhere,<sup>1, 2</sup> but it is relevant to report on the type of measurements made of the stuttering symptom and their relationships to each other. This latter aspect can be considered independently of any drug effects on the subjects who participated in the study.

**Criterion of What Constitutes Stuttering.** The construct and application of measurements and ratings of stuttering must take into account a criterion of what constitutes stuttering. This becomes particularly important when differentiations must be made between the non-fluencies that can be detected in normal speakers and stutterers. Johnson has observed that “interjections, revisions, and phrase repetitions” can sometimes be considered as “normal” disfluencies.<sup>7</sup> Boehmler had indicated that two groups of judges (trained and untrained) classified part-word repetitions as “stuttering” as

frequently for stutterers as for non-stutterers.<sup>4</sup> The judges agreed more when stuttering was severe and when "normal" non-fluencies were mild. Williams and Kent found that syllable repetitions and prolongations were more consistently identified as "stuttering" by judges.<sup>21</sup>

Johnson listed eight features of speech which he thought to be representative of disfluencies and this became known as the Iowa Speech Disfluency Test.<sup>7</sup> Young used a modified version of this test, and considered five rather than the eight features, in view of the rarity of occurrence of "phrase repetitions, incomplete phrases and broken words."<sup>23</sup> His categories were:

- (a) Interjections.
- (b) Part-word repetitions.
- (c) Word-phrase repetitions (Johnson considered these as two separate features).
- (d) Prolongations, including broken words (Johnson considered these as two separate features).
- (e) Revisions, including incomplete phrases (Johnson considered these as two separate features).

Young then used multiple correlation procedures to analyse the speech samples of stutterers as rated by listeners.<sup>23</sup> As a result of these procedures he modified his categories further and this was the criterion adopted in the study under discussion. The categories were:

- (a) Syllable and sound repetitions.
- (b) Sound prolongations.
- (c) Broken words (or broken utterances).
- (d) Words involving apparent unusual stress or tension.

In addition to the above, the writer added word repetitions with "syllable and sound repetitions" as she feels that this feature occurs fairly frequently, especially on short words. The criterion followed appears feasible to use as a classification for denoting stuttering moments and rating the severity of stuttering.

**Severity Ratings of Stuttering.** Various measurements can be applied to ascertain the severity of stuttering. Physiological changes can be noted such as alterations in breathing, heart rate, electrical skin response, etc. The duration of the stuttering moments can also be measured. In a personal communication, Van Riper pointed out that there is still no adequate way of measuring the severity of stuttering.<sup>19</sup> He felt that the problem lies in the fact that stuttering is a compound of fear and struggle. He considered the use of physiological measurements but suggested at the same time that these were subject to too many variables and would differ from subject to subject according to the nature of their symptoms. The use of physiological measurements, although theoretically plausible, were, for practical reasons, rejected for the purposes of this study.

One satisfactory measurement of the severity of stuttering which has received some attention over recent years, is the use of an "equal-appearing intervals scale" to rate severity based on audible characteristics. The use of such a scale represents a classic psychophysical method. The subject (or experimenter) sorts a number of stimuli into specified categories where the intervals appear to the experimenter to be equal.

Sherman has done a great deal of work on severity scales, and together with Lewis, she first applied a nine-point scale consisting of equal-appearing intervals where 1 represented "least severe stuttering" and 9 represented

“most severe stuttering.”<sup>15</sup> Fifteen judges rated speech samples on this scale and their judgements were found to be satisfactorily reliable. There was evidence that this severity scale had internal consistency. This scale was applied to permanent records (tape recordings) of the audible characteristics of stuttering. Sherman then used the same scale to ascertain if observers could make reliable judgements of continuous stuttered speech.<sup>13</sup> She demonstrated that this method is reliable, and is experimentally and clinically useful for assigning a rank order position of severity.

Johnson suggests a seven-point equal-appearing intervals scale where he presents a description of the type of stuttering that each rating represents.<sup>8</sup> Cullinan *et al.* compared five-, seven-, and nine-point equal-appearing intervals scales and found that the scales were significantly correlated and that reliable values could be obtained from any one of these.<sup>5</sup>

Apart from measures of the severity of stuttering, the frequency with which stuttering occurs must contribute to the listeners' evaluation of the severity of the symptom. This aspect was studied by Sherman and Trotter who questioned the relationship between the frequency count of the moments of stuttering and measures of severity made on an equal-appearing intervals scale.<sup>16</sup> From the data presented by eleven observers, they found that these two aspects were significantly and positively correlated, but they do point out that the relationship between the two measures is not strong enough to allow one measure to predict the other. Sherman *et al.* compared three measures of stuttering severity:

- (a) Reading time.
- (b) Frequency moments of stuttering.
- (c) Scale values from listeners' ratings.<sup>17</sup>

All interrelationships were found to be statistically significant and the strength of the relationship was highest between the frequency counts and the rated severity.

The visual characteristics of stuttering may be as important to measure as the audible characteristics when making severity ratings. It is difficult, however, to obtain such information for permanent records due to the expense of filming. It is necessary to establish, nevertheless, how close or how far severity ratings based on audible characteristics differ from those based on visual and audible features. In this regard a detailed study was conducted by Williams *et al.*, who studied the ratings of stuttering by audio, visual and audiovisual cues making use of synchronized photography and sound recording.<sup>22</sup> They found that the frequency count of stuttering and scale values representing severity are more reliable when obtained by audio and audiovisual observation than when obtained by visual observation. They suggested that the use of audio cues alone is sufficient for obtaining useful and reliable measures of the frequency and severity of stuttering. It is possible, the present writer feels, to consider the relative importance and influence of visual cues when marking the severity ratings of stuttering during the stutterers' act of reading or speaking, where audible cues would also be present apart from the visual characteristics. If the speech is recorded and played back at a later stage when the memory of the visual cues ceases to be influential, comparative data of

ratings could indicate whether visual cues added significantly to the initial ratings of severity.

Johnson has said that the advantage of a severity rating scale with stutterers ensures a certain degree of uniformity and comparability of judgements.<sup>8</sup> He cautions that such a rating can have its limitations and that clinical judgements of stutterers are still important.

A relevant point that can be considered here is the reliability of one observer to make severity ratings of stuttering. Cullinan *et al.* found that judges tend to agree better with themselves than with other judges.<sup>5</sup> Their evidence suggests that it is not sufficiently reliable to use a single judge on a single rating scale. It would seem feasible and convenient to employ a single observer to make severity ratings which can be compared with ratings made by a group of judges. This should indicate the kind of agreement among the judges and the reliability of the single observer's judgements. Sherman studied this aspect by comparing individual observers and found, in contrast to Cullinan *et al.*, that reliable scale values can be obtained from a single observer.<sup>14</sup>

**Frequency of Stuttering Moments.** The frequency with which stuttering occurs contributes to the severity of the stuttering symptom. Like severity, frequency of stuttering can increase or decrease, depending on the tension and anxiety felt by the stutterer in communicative situations. The counting of stuttering moments constitutes a basic measurement of the disorder. It is customary to have a subject read a passage and to mark the words stuttered on. It is also possible to count the frequency of stuttering during spontaneous speech where tape recordings can be transcribed and the moments marked and studied. As with the severity ratings of stuttering, visual cues exhibited by the stutterer can influence the experimenter's markings of the moments of stuttering while the subject is present and reading.

Two important dimensions of stuttering behaviour are calculated on the basis of the frequency count—adaptation and the consistency effect.

*Adaptation.* This is a well-known phenomenon in stuttering where stuttering moments tend to decrease progressively with successive oral readings of the same passage. There appear to be individual differences, however, where some stutterers do not show any adaptation or may even produce increased stuttering with every reading.<sup>8</sup> Newman considered the possibility of classifying stutterers into groups on the basis of their adaptation performances.<sup>10</sup> He re-examined data on adaptation and found that there were instances where subjects did not reveal any adaptation. The first study to note the phenomenon of adaptation in stuttering was conducted by Van Riper and Hull in 1934.<sup>20</sup> They attributed the decrease in the stuttering moments to the stutterer's subjective adaptation, either to the situation or to the reading material. Since then many studies have been published on this aspect and the general attitude appears to be that the stutterers experience less anxiety with each reading, thereby stuttering less.

*Consistency.* Consistency effect of stuttering was first noted by Johnson and Knott who found a significant tendency for subjects to stutter on the same words from reading to reading of the same passage.<sup>9</sup> Johnson *et al.* feel that the importance of knowing the consistency in stuttering is that it may indicate

how strongly stuttering responses are associated with stimuli or cues that have been conditioned.<sup>8</sup> The anticipation and expectancy of stuttering is closely related to the consistency effect.

*Reading Rate.* It has been commonly observed that stutterers take longer to read than non-stutterers. This time factor contributes to the severity of the problem. Sander reported that reading rate was a highly stable measure of stuttering severity.<sup>12</sup> He considered the relationship between reading rate and the frequency moments of stuttering, and concluded that the two measurements can be used as a tool for evaluating the speech improvement of stutterers. The computation of the reading rate in words per minute is outlined by Johnson *et al.*<sup>8</sup> The number of words contained in a reading passage is divided by the number of seconds taken to read it and this is multiplied by sixty to convert the final rate to words per minute. This measurement, requiring only the use of a stop-watch, is perhaps one of the most objective methods of assessing any change in stuttering. The severity of the stuttering symptom, together with the frequency of its occurrence, will directly influence the rate of reading.

### Procedure in Present Study

The study extended over fifteen weeks which were divided into five periods of three weeks each. Forty-six subjects participated—37 were Europeans (mean age: 26 years 6 months) and 9 were Africans (mean age: 20 years 1 month). A number of testing procedures was administered to subjects individually once in three weeks. The purpose of these three-weekly assessment-interviews was to collect data pertaining to various aspects of the subjects' stuttering. Apart from the measurements of aspects of stuttering, the subjects also completed a daily questionnaire for the duration of the experiment and made verbatim reports, but results of these are not pertinent to the subject matter under consideration.

Although the measurement of aspects of stuttering was done with a view to ascertaining the efficacy of medication, the measurements can be studied in isolation without any spurious effect from the drug. The purpose of this present report is to indicate the inter- and intra-relationships between the measurements themselves.

### MATERIAL

The reading passage utilized was a slight modification of a portion of the "Rainbow" passage by Fairbanks.<sup>6</sup> The passage was modified so that it could be divided into two sections—the first part consisting of 200 words, and the second part of 110 words. During the first three assessment periods only the 200 word passage was presented, and during the last two assessment periods the complete 310 word passage was given. The extra 110 word passage was included in order to ascertain whether any adaptation to the same passage (first 200 words) had occurred. Thus the 110 word passage, as the non-adaptation passage, acted as a control.

**1. Experimenter's Severity Rating of Stuttering on Subjects' Reading.** The rating of the severity of stuttering was made by the experimenter after the subject had read the passage.

(a) The rating was made on a nine-point equal-appearing intervals scale where "1" represented no stuttering, "9" very severe stuttering, and "5" represented average stuttering. The other points on this scale represented values falling between these points. The rating was made by the experimenter immediately after the subject had completed his reading.

(b) A second rating on the same scale as described above was made when the tape-recording of the reading was played back. This was done soon after the subject left the experimenter and it was felt necessary since it was possible that visual cues displayed by the subjects during the reading might have affected the first severity rating made:

(c) A third severity rating was made, without reference to the above two ratings, using the same scale. This was made when the tape-recording was played back approximately 10 days after the initial reading. This third rating was based predominantly on auditory characteristics. It was also considered a crucial one in that samples of tape-recordings were to be played to judges who would react only to these auditory cues, thus making comparative data possible.

**2. Experimenter's Severity Rating of Stuttering on Subjects' Spontaneous Conversation.** This rating was made on the same nine-point equal-appearing intervals scale described above. Spontaneous conversation took place when the subjects first arrived for the interview. At the end of the interview, if the information was not volunteered in enough detail, the subjects were asked to comment on their speech, on the study, etc. This rating was made separately from the readings as it was felt that with many stutterers, there is a marked disparity between the amount of stuttering in reading and in spontaneous speech performances. No attempt was made to structure the conversation and no tape recording was made.

**3. Subjects' Severity Rating on Their own Stuttering in Reading.** After a subject had read the passage he was asked to rate himself on the same scale used in the above measurements. The instruction to the subject was as follows:

*On the following scale, rate your stuttering on the passage you have just read.*

1	2	3	4	5	6	7	8	9
No				Average				Very Severe
Stuttering				Stuttering				Stuttering

**4. Subjects' Severity Rating of Their own Stuttering on a Playback of the Recording of the Reading.** Approximately thirty minutes after the subject had recorded the reading passage, he was asked to hear the tape recording and to rate himself on the same scale. The instruction to the subject was:

*The reading you have just made will be played back to you. Listen to it and rate your stuttering on the following scale:*

The scale was presented in the same manner as in 3 above. It was felt that the confrontation of hearing their speech might change their judgement of the severity which they had previously made on their reading. It was expected that, due to poor attitudes towards speech, the subjects would make a slightly higher rating than previously. This information could be of value for comparative purposes, in terms of self-judgements of the severity of stuttering based on the same material judged under different conditions.

**5. Frequency Count of Stuttering Moments during the Subjects' Reading.** While the subject read, the experimenter marked on a copy of the reading passage the words on which stuttering occurred. During the last two periods of the study, when the additional 110 word passage was introduced together with the familiar 200 word passage, the frequency count was taken separately for both passages.

The tape-recording of the reading was played back later during the same day of the recording. A second frequency count was taken and no reference was made to the markings on the first copy of the passage. This aspect was considered twice since it is possible that visual cues representing struggle responses, especially those associated with non-vocalized blocks, would be noticed during the reading when the subject was present. It can be expected, therefore, that the frequency of stuttering is greater when taken during reading as compared with the count taken on the playback of the recording.

**6. Reading Rate Measured in Words per Minute.** When the tape-recordings were played back during the same day of the readings, they were timed with a stop-watch. The reading rate was calculated according to the number of words contained in the passage (either 200 or 310 words). As discussed previously, reading rate is probably the most objective dimension of stuttering behaviour, since it does not rely on any subjective evaluation made by an observer. The amount of stuttering moments and the severity with which stuttering occurs, will directly influence the time it takes a stutterer to read.

**7. Subjects' Severity Rating of Stuttering for the Same Day, Prior to the Assessment-Interview.** The subjects were asked to rate their stuttering as it was for the same day, prior to the interview. The question was put to them in this way:

*What was your speech like today before coming here? Was it the same as usual, slightly better, much better, slightly worse, much worse?*

This question was very familiar to them as it followed the same format as the daily schedule questions which they were completing. Their answer was rated by the experimenter on the following scale:

1	2	3	4	5
Much Better	Slightly Better	Same	Slightly Worse	Much Worse

This rating was made as it was felt that the interview situation might have been perceived as being tense by some subjects, and the likelihood might have arisen that their speech symptoms, during the interview, might have been worse than the speech produced during the day, before coming to the clinic. At the same time the subjects were asked to comment on the "usualness" or "unusualness" of the day they had just experienced. This question was asked in order to elicit information as to whether anything untoward had occurred that could possibly affect their speech.

*Reliability of Experimenter's Severity Ratings.* A month after the completion of the experiment, five judges rated the severity of stuttering from the recordings. The purpose of this was to test the reliability of the experimenter's ratings and the degree of agreement with judges. The judges were qualified speech therapists. Recordings representing any two of the five assessment-

interviews for each subject were selected according to a set of random numbers and totalled ninety-two samples. These recordings were removed from the original tape reels, spliced together and placed, according to the randomized order, onto large reels. The instructions to the judges included the following:

*The tape samples you will hear have been made by persons who regard themselves as stutters. In each sample the "Rainbow" passage by Fairbanks is read. In some instances a 200 word portion of the passage will be read, and in others a 310 word portion will be read.*

*You are requested to make a rating of the severity of stuttering for each sample.*

*This rating is to be made on a nine-point equal-appearing intervals scale on which a rating of 1 means "no stuttering", a rating of 2 means "very mild stuttering" and a rating of 9 means "very severe stuttering". A rating of 5 indicates "average severity". The other values on the scale represent equal intervals between 1 and 9. Please give only one rating for each sample heard.*

The judges were requested to consider the same criterion of stuttering speech as did the experimenter.

## Results and Discussion

**Comparison of the Experimenter's Severity Ratings of Stuttering with those Made by Five Judges.** Judges rated any two recordings of the reading made by each subject out of a possible five. The judges' ratings were compared with each other and with the experimenter's.

The results were inter-correlated where the ratings of the five judges were used, together with the experimenter's three ratings (factors 1a, 1b and 1c referred to above). The correlation matrices depict the judges' and experimenter's ratings for the first and second recordings chosen, and the correlations between the ratings of the two recordings. The inter-correlations were derived from raw scores, means and standard deviations.

TABLE 1: INTER-CORRELATIONS OF THE SEVERITY RATINGS OF STUTTERING MADE BY FIVE JUDGES AND THE EXPERIMENTER FOR THE FIRST GROUP OF RECORDINGS HEARD

1	2	Judges			Experimenter		
		3	4	5	E1	E2	E3
1.00							
.93	1.00						
.85	.91	1.00					
.86	.92	.92	1.00				
.92	.94	.91	.90	1.00			
.88	.88	.85	.88	.88	1.00		
.90	.92	.88	.89	.91	.96	1.00	
.89	.91	.88	.88	.90	.95	.97	1.00

Significance at 1% level = .36

The matrices show high correlations for the experimenter's three severity ratings (reading, first and second playback). The correlations range from .95 to .97 for both recordings (Table 1 and 2). The correlations between the experimenter's ratings and those of the judges ranged from .85 to .92 for the first recordings heard and from .86 to .93 for the second recordings heard. These are higher than the correlations *between* the two recordings for the individual judges (inter-judge), where they ranged from .71 to .87 (Table 3),

TABLE 2: INTER-CORRELATIONS OF THE SEVERITY RATINGS OF STUTTERING MADE BY FIVE JUDGES AND THE EXPERIMENTER FOR THE SECOND GROUP OF RECORDINGS HEARD

1	2	Judges			Experimenter		
		3	4	5	E1	E2	E3
1.00							
.92	1.00						
.85	.91	1.00					
.92	.95	.92	1.00				
.93	.91	.90	.93	1.00			
.87	.92	.89	.93	.90	1.00		
.88	.92	.88	.93	.90	.95	1.00	
.86	.91	.86	.92	.90	.96	.96	1.00

Significance at 1% level — .36

TABLE 3: INTER-CORRELATIONS OF THE SEVERITY RATINGS OF STUTTERING MADE BY FIVE JUDGES AND THE EXPERIMENTER BETWEEN THE TWO GROUPS OF RECORDINGS HEARD

Second Recording	First Recordings								
		Judges					Experimenter		
		1	2	3	4	5	E1	E2	E3
Judges	1	.82	.82	.78	.72	.81	.76	.80	.80
	2	.83	.87	.80	.78	.84	.81	.85	.85
	3	.79	.86	.85	.76	.85	.80	.83	.84
	4	.81	.85	.80	.77	.83	.82	.84	.84
	5	.76	.81	.79	.71	.82	.77	.81	.82
Exp.	E1	.73	.79	.73	.67	.75	.75	.79	.80
	E2	.72	.75	.72	.67	.74	.76	.80	.80
	E3	.72	.74	.68	.62	.73	.72	.76	.77

Significance at 1% level = .36

or for the *between-recordings* for the experimenter, where correlations ranged between .75 to .80 (Table 3). This last correlation is possibly affected by actual differences in the subjects' stuttering from one period to another—an effect by which the judges could not have been influenced. The correlations between the judges themselves (*intra-judge*) are all high and show a satisfactory measure of agreement. These findings support the evidence contributed by Cullinan *et al.*, where *intra-judge* reliability coefficients tended to be greater than the *inter-judge* reliability coefficients, i.e. the judges tend to agree better with themselves than with other judges.<sup>5</sup>

From the matrices it can be seen that the experimenter obtained very high correlations between her *three ratings* for each recording—the correlations for the first recording ranges from .95 to .97 and for the second from .95 to .96. This would indicate that there is little difference between the ratings of the severity of stuttering during the reading while the subject was present and a rating made ten days later from a recording. It appears then that, in this study, the visual cues displayed by the subjects while reading had very little effect on the severity ratings made. It can be postulated that tensions displayed during the actual reading (where stuttering is physically manifest and can be “seen”)

are reflected in the manner of reading itself so that they are auditorily detected in the recordings heard.

From the results it would appear that the experimenter's ratings are highly correlated with each other and are satisfactorily in agreement with the five judges. These ratings can therefore be considered as a reliable measurement of the severity of stuttering. This finding also supports Sherman's observation that a single observer is sufficient to make reliable ratings of the severity of stuttering.<sup>14</sup>

#### COMPARISON OF STUTTERING MEASUREMENTS OVER FIVE PERIODS

A correlation of 10 factors (points 1a, 1b, 1c, 2, 3, 4, 5a, 5b, 6, 7, discussed above) to observe the nature of the relationship between the measurements themselves was carried out. The matrices for each period are presented in Tables 4-8. The means and standard deviations for the respective periods are included alongside the relevant matrix. The IBM 704, 8K Computer of the C.S.I.R., Pretoria, was used in order to derive these matrices. Although it may have been more convenient to refer to one table representing the five periods, the exact nature of the correlations would not have been evident as the scores would have differed due to trends over the five periods. Partial correlations were calculated for each period between the experimenter's severity rating (factor 1c) and reading rate (factor 6) eliminating any joint correlation with the frequency count of stuttering moments. This was done in order to establish whether the reading rate or the frequency count influenced the severity ratings made by the experimenter.

**Experimenter's Three Severity Ratings.** The three ratings are highly correlated extending from .93 to .98. There is a tendency for the correlations to increase from the first period to the fifth period. There is a steady decline in the means from the first to the fourth period (.48 to .42), and a decline in the standard deviations from 2.0 to 1.5 which indicates that the spread of the scores narrowed. These results may indicate progressive facility with the use of the scale on the part of the experimenter and/or familiarity of the subjects with the experimental situation. On the other hand the means and the standard deviations are increased during the fifth period. The increased standard deviations are probably due to the greater variability between the subjects during the last period of the study. The increased mean scores may also be due to this variability within the subjects, or to a worsening in the stuttering condition as the medication periods had passed.

The severity ratings correlate very highly with *reading rate* ranging from .79 to .92. The correlations are also high with the two *frequency counts* which range from .46 to .71. The range of correlations of the experimenter's ratings with the rating made for *conversation* declines steadily from .71 to .46—the correlations, however, are still significant (significance = .29). The experimenter's severity ratings correlates comparatively poorly with the *subjects' severity ratings* of their own speech—ranging from .37 to .70. This is to be expected as the standards used by the experimenter and by the subjects must differ due to the element of subjectivity present on the part of the stutterers.

TABLE 4: CORRELATION MATRIX INDICATING THE RELATIONSHIP BETWEEN TEN FACTORS FOR THE FIRST PERIOD OF THE STUDY FOR FORTY-SIX SUBJECTS  
THE MEAN SCORES AND STANDARD DEVIATIONS FOR EACH FACTOR ARE INCLUDED

Factors	Experimenter's Severity Ratings			Conversation	Subjects' Severity Ratings		Frequency Counts		Reading Rate	Subjects' Rating for Day
	Reading 1a	1st Pl.b. 1b	2nd Pl.b. 1c	2	Reading 3	1st Pl.b. 4	Reading 5a	1st Pl.b. 5b	6	7
1a	1.000									
1b	0.933	1.000								
1c	0.930	0.937	1.000							
2	0.696	0.651	0.660	1.000						
3	0.638	0.604	0.658	0.234	1.000					
4	0.506	0.508	0.521	0.228	0.496	1.000				
5a	0.566	0.642	0.618	0.493	0.375	0.141	1.000			
5b	0.552	0.628	0.608	0.480	0.380	0.111	0.995	1.000		
6*	-0.891	-0.880	-0.922	-0.607	-0.653	-0.462	-0.722	-0.704	1.000	
7	-0.002	-0.025	0.004	0.222	0.040	-0.121	0.074	0.086	0.024	1.000

Significance level for correlations at 5% level = .29

\*Partial Correlations between experimenter's severity rating (1c) and reading rate, excluding joint correlation with frequency count = .88

Factors	Mean	Std. Dev.
1a	4.891	1.853
1b	4.804	1.916
1c	4.848	1.988
2	4.565	1.996
3	5.935	2.081
4	6.587	2.082
5a	44.196	37.529
5b	43.587	37.834
6	92.380	46.682
7	3.174	0.677

TABLE 5: CORRELATION MATRIX INDICATING THE RELATIONSHIP BETWEEN TEN FACTORS FOR THE SECOND PERIOD OF THE STUDY FOR FORTY-SIX SUBJECTS  
THE MEAN SCORES AND STANDARD DEVIATIONS FOR EACH FACTOR ARE INCLUDED

Factors	Experimenter's Severity Ratings			Conversation	Subjects' Severity Ratings		Frequency Counts		Reading Rate	Subjects' Rating for Day
	Reading Ia	1st Pl.b. Ib	2nd Pl.b. Ic	2	Reading 3	1st Pl.b. 4	Reading 5a	1st Pl.b. 5b	6	7
Ia	1.000									
Ib	0.971	1.000								
Ic	0.943	0.964	1.000							
2	0.695	0.673	0.644	1.000						
3	0.501	0.536	0.525	0.207	1.000					
4	0.454	0.494	0.476	0.247	0.707	1.000				
5a	0.639	0.681	0.648	0.375	0.516	0.236	1.000			
5b	0.641	0.685	0.652	0.383	0.516	0.259	0.998	1.000		
6*	-0.845	-0.862	-0.863	-0.654	-0.525	-0.401	-0.761	-0.763	1.000	
7	-0.317	-0.334	-0.388	-0.200	0.105	-0.173	-0.132	-0.139	0.322	1.000

Significance level for correlations at 5% level = .29

\*Partial Correlations between experimenter's severity rating (Ic) and reading rate, excluding joint correlation with frequency count = -.75

Factors	Mean	Std. Dev.
Ia	4.674	1.851
Ib	4.609	1.757
Ic	4.652	1.754
2	4.783	1.861
3	5.457	2.105
4	5.522	2.137
5a	39.891	39.143
5b	39.522	38.172
6	100.596	49.443
7	2.674	0.790

TABLE 6: CORRELATION MATRIX INDICATING THE RELATIONSHIP BETWEEN TEN FACTORS FOR THE THIRD PERIOD OF THE STUDY FOR FORTY-SIX SUBJECTS  
THE MEAN SCORES AND STANDARD DEVIATIONS FOR EACH FACTOR ARE INCLUDED

Factors	Experimenter's Severity Ratings			Conversation	Subjects' Severity Ratings		Frequency Counts		Reading Rate	Subjects' Rating for Day
	Reading 1a	1st Pl.b. 1b	2nd Pl.b. 1c	2	Reading 3	1st Pl.b. 4	Reading 5a	1st Pl.b. 5b	6	7
1a	1.000									
1b	0.969	1.000								
1c	0.956	0.984	1.000							
2	0.562	0.540	0.550	1.000						
3	0.686	0.737	0.697	0.218	1.000					
4	0.596	0.592	0.563	0.257	0.780	1.000				
5a	0.652	0.693	0.707	0.405	0.517	0.243	1.000			
5b	0.656	0.696	0.711	0.403	0.518	0.249	0.999	1.000		
6*	-0.883	-0.879	-0.900	-0.632	-0.627	-0.526	-0.736	-0.738	1.000	
7	0.038	0.038	0.020	0.025	0.211	0.050	0.071	0.057	0.005	1.000

Significance level for correlations at 5% level = .29

\*Partial Correlations between experimenter's severity rating (1c) and reading rate, excluding joint correlation with frequency count — — .79

Factors	Mean	Std. Dev.
1a	4.435	1.747
1b	4.457	1.798
1c	4.391	1.732
2	4.848	1.813
3	5.478	2.383
4	5.413	2.427
5a	33.783	34.207
5b	33.674	34.172
6	106.276	50.690
7	2.935	0.772

Journal of the South African Logopedic Society, Vol. 14, No. 1: September 1967

TABLE 7: CORRELATION MATRIX INDICATING THE RELATIONSHIP BETWEEN TEN FACTORS FOR THE FOURTH PERIOD OF THE STUDY FOR FORTY-SIX SUBJECTS  
THE MEAN SCORES AND STANDARD DEVIATIONS FOR EACH FACTOR ARE INCLUDED

Factors	Experimenter's Severity Ratings			Conversation 2	Subjects' Severity Ratings			Frequency Counts		Reading Rate 6	Subjects' Rating for Day 7
	Reading 1a	1st Pl.b. 1b	2nd Pl.b. 1c		Reading 3	1st Pl.b. 4	Reading 5a	1st Pl.b. 5b			
1a	1.000										
1b	0.961	1.000									
1c	0.945	0.969	1.000								
2	0.519	0.566	0.498	1.000							
3	0.531	0.564	0.612	0.028	1.000						
4	0.460	0.441	0.472	-0.028	0.714	1.000					
5a	0.461	0.565	0.592	0.336	0.520	0.155	1.000				
5b	0.455	0.558	0.587	0.332	0.520	0.148	1.000	1.000			
6*	-0.785	-0.831	-0.827	-0.467	-0.532	-0.360	-0.638	-0.637	1.000		
7	-0.084	-0.066	-0.062	0.181	0.125	0.073	-0.029	-0.033	0.107	1.000	

Significance level for correlation at 5% level = .29

\*Partial Correlations between experimenter's severity rating (1c) and reading rate, excluding joint correlation with frequency count = -.73

Factors	Mean	Std. Dev.
1a	4.217	1.489
1b	4.283	1.587
1c	4.152	1.460
2	4.196	1.668
3	5.130	2.267
4	5.174	2.184
5a	25.239	28.678
5b	25.109	28.828
6	104.883	38.285
7	2.761	0.822

Myrtle L. Aron

TABLE 8: CORRELATION MATRIX INDICATING THE RELATIONSHIP BETWEEN TEN FACTORS FOR THE FIFTH PERIOD OF THE STUDY FOR FORTY-SIX SUBJECTS  
THE MEAN SCORES AND STANDARD DEVIATIONS FOR EACH FACTOR ARE INCLUDED

Factors	Experimenter's Severity Ratings			Conversation	Subjects' Severity Ratings			Frequency Counts		Reading Rate	Subjects' Rating for Day
	Reading 1a	1st Pl.b. 1b	2nd Pl.b. 1c	2	Reading 3	1st Pl.b. 4	Reading 5a	1st Pl.b. 5b	6	7	
1a	1.000										
1b	0.974	1.000									
1c	0.944	0.967	1.000								
2	0.463	0.444	0.446	1.000							
3	0.513	0.552	0.574	-0.097	1.000						
4	0.370	0.393	0.419	-0.120	0.761	1.000					
5a	0.565	0.570	0.556	0.172	0.456	0.104	1.000				
5b	0.563	0.568	0.554	0.172	0.458	0.109	1.000	1.000			
6*	-0.812	-0.796	-0.820	-0.427	-0.458	-0.224	-0.641	-0.638	1.000		
7	0.245	0.227	0.212	0.229	0.321	0.324	0.079	0.075	-0.110	1.000	

Significance level for correlations at 5% level = .29

\*Partial Correlations between experimenter's severity rating (1c) and reading rate, excluding joint correlation with frequency count = -.73

Factors	Mean	Std. Dev.
1a	4.174	1.596
1b	4.217	1.604
1c	4.370	1.638
2	4.522	1.709
3	4.870	2.391
4	4.870	2.372
5a	25.500	32.833
5b	25.543	33.151
6	113.027	41.887
7	2.935	0.712

The *partial* correlations calculated for each period between the experimenter's severity rating and the reading rate, excluding the joint correlation with the frequency count, indicate that the actual speed of reading is an important element in the severity ratings made. The partial correlations are all high extending from  $\cdot 72$  to  $\cdot 88$ .

Thus it appears that the experimenter's severity ratings are most highly correlated with the reading rate and then with the frequency counts of stuttering moments. These latter two factors are regarded in the literature as being fairly accurate measures of stuttering behaviour, and the satisfactory correlation with the experimenter's severity ratings indicates the relative closeness between these three forms of stuttering measurement.

**Experimenter's Severity Ratings of Stuttering based on Conversation.** As pointed out the correlations between the ratings based on conversation and on the reading of the passage decline steadily to the fourth period, whereas the means of ratings based on conversation show no such trend—they rise from  $4\cdot 6$  in the first period to  $4\cdot 8$  in the second and third periods, falling to  $4\cdot 2$  in the fourth period, and then rise to  $4\cdot 5$  in the fifth period. This was not accompanied by any increase in the standard deviations—on the contrary, they declined from  $2\cdot 0$  to  $1\cdot 7$ .

It is possible, that because no set amount of conversation (or topic) was required from each subject, the choice of topic, the manner of expression, and the amount spoken might have influenced the experimenter from period to period. In this regard it would obviously have been better to set the subjects a particular topic of conversation during each period, which could be recorded, transcribed from the recordings, and then rated for severity. This is what Johnson did with his "Job" task where the subjects were asked to talk about aspects of their employment.<sup>7</sup>

**Subjects' Severity Ratings of Their own Speech on Reading.** The correlations between the subjects' ratings of their speech on reading and that based on the recording for the five periods are:  $\cdot 49$ ,  $\cdot 71$ ,  $\cdot 78$ ,  $\cdot 71$ ,  $\cdot 76$  (all significant). The low correlation in the first period was probably due to the fact that that the subjects were not at that stage familiar with the routine of the experiment and rating scale. The correlations of the subjects' ratings and the experimenter's ratings of severity are also comparatively low with a wide range from  $\cdot 37$  to  $\cdot 70$ . The role of subjectivity and personal feelings about stuttering cannot be discounted. The correlation with *reading rate* is low, extending from  $\cdot 22$  (not significant) to  $\cdot 65$ . An interesting observation is that the correlations of the subjects' ratings based on the recording (factor 4) with that of reading rate, is always lower than the correlation between the subjects' ratings on their original reading and reading rate. This indicates that the stutterers rated subjectively and were probably influenced by their feelings about their speech, especially when hearing the recordings. The fact that reading rate is an objective measure indicates that the stutterer's feelings about his own performance should be viewed as a separate entity when comparing ratings and assessments of stuttering behaviour. The subjects' severity ratings of their speech on reading and on the recording correlates very poorly with their own ratings

for their speech *for the same day* prior to the interview—they range from  $\cdot 04$  to  $\cdot 32$ . These results are not surprising since the subjects' speech on reading is a situation not practically comparable with spontaneous speech which could be worse or better than while reading. Reading does not constitute "spontaneous propositionality," and further the reading passage was chosen for its lack of emotional content, i.e. reading and spontaneous speech represent two different forms of communication.

**Frequency Counts of the Moments of Stuttering.** The correlations between the two counts taken on the frequency of stuttered moments are extremely high, ranging from  $\cdot 995$  to  $1\cdot 00$ . This indicates that there was little discrepancy between the two counts taken for each subject, which implies that any visual cues present in the first rating were at a minimum and had little influence on the frequency of the stuttered moments counted. The correlation between the frequency counts and *reading rate* rose from  $-\cdot 70$  to  $-\cdot 76$  between the first and second periods, and then declined steadily to  $-\cdot 64$  in the fifth period. This was accompanied by a steady fall in the mean frequency counts from 43 to 25 over the five periods. The reading rate is affected by both the frequency of stuttering and the length of the blocks, and the frequency counts by the number of blocks only. The decreasing correlations, together with the fact that the mean number of frequency counts had fallen to a little more than half the original value, may suggest that the shorter blocks tended to be eliminated. Following on from this it can be interpreted that the reading rate, towards the end of the experiment, tended to be influenced more by the length of the blocks than by the frequency with which they occurred. This is corroborated by the decline in the *partial* correlation between the experimenter's severity ratings and reading rate, where the joint correlation with the frequency counts was eliminated—the range was from  $\cdot 88$  to  $\cdot 72$  over the five periods.

**Reading Rate.** This factor has already been discussed in terms of its correlation with the above eight factors. The *partial* correlations carried out for each period between the experimenter's severity rating (factor 1c) and the reading rate (where the joint correlation with the frequency count was eliminated) indicate that the speed of reading influenced the experimenter's severity ratings to a large extent in the first period, but they declined towards the fifth period. The correlations are all high and extend from  $\cdot 88$  to  $\cdot 72$ .

**Subject's Severity Rating for Their Speech for the Same Day Prior to the Interview.** This rating bears little relationship to the measurements taken. Its inclusion was only to provide a rough indication of whether the ratings made on conversation during the interview corresponded with the subjects' report of their stuttering during the same day of the interview. No relationship was found—the correlations range from  $\cdot 03$  to  $\cdot 23$ , none of which is significant. The form of speech exhibited in the clinic situation can be expected to differ from the type of speech produced outside this situation. In addition many extraneous factors could have influenced the subjects to make different assessments from the experimenter, e.g. the stutterers' subjective attitude towards speech, the experimenter's practice in making severity ratings.

## Conclusion

It was found in this study that a single investigator can make valid ratings of the severity of stuttering. This was supported by the high and significant correlations between five judges and the experimenter. The extremely high correlations between the experimenter's three severity ratings of the same reading for individual subjects taken at different times, would seem to indicate that visual cues displayed by stutterers while reading had little effect on the severity ratings made. It was postulated that tensions, where physical manifestations accompanied stuttering, are reflected in the manner of speech itself, so that it can be detected from tape recordings.

Frequency of the occurrence of stuttered moments may not necessarily be an adequate measure of stuttering change and severity. Severity ratings of stuttering correlated most highly with reading rate. The frequency counts of the moments of stuttering also correlated well, but not as highly. The correlations between the frequency count and reading rate extended from .64 to .76. From the results it appeared that the reading rate, towards the end of the experimental period, tended to be influenced more by the length of the blocks than by the frequency with which they occurred. It is postulated that if the severity of stuttering undergoes change within the same individual over a period of time, this would be corroborated by change in reading rate, rather than frequency. Thus it appears that frequency counts of stuttering, while a fairly reliable measurement in itself, is not a sufficiently reliable measure of stuttering severity when used alone.

The frequency with which stuttering occurs is referred to in the literature as representing one measure of the *severity* of stuttering.<sup>3, 8, 23</sup> It can be generalized from the results here that we should differentiate more clearly between the terms "frequency" and "severity" of stuttering. There is no doubt that the amount of frequency of stuttering will contribute to an impression of its severity, but it would appear that frequency as such should not be regarded as an independent and accurate measure of severity.

Reading rate, while being an objective measure of stuttering, also correlates highly with severity ratings, and therefore indicates that it is a reliable measure which can be taken of stuttering. However, it must be pointed out that this refers to reading only. To rely on rate of speech from transcriptions of spontaneous speech would be difficult as so many variables can confound the issue, e.g. the natural manner and fluency of expression apart from the stuttering moments, the content of what is expressed. In addition, some comparison of speaking rates with normal speakers using spontaneous speech would have to be considered.

The clinical impression that stutterers are poor judges of their own speech performance was borne out by the results of this study. This implies that we should accept that different standards would effect the assessment of stuttering behaviour as compared to the assessment made by stutterers themselves. It would be relevant for therapists to consider their own rating of stuttering and what constitutes improvement, and separately the stutterers should make their own ratings as to what improvement they feel they are making. It is clear that

we cannot equate both forms of assessment. While there is still a great deal to learn about stuttering behaviour, there is an indication that we can go some way now in attempting to assess aspects of it which have a relatively good measure of reliability.

### Summary

The report is concerned with the types of measurements that can be made of stuttering behaviour and their relationships to each other. A review of the different measurements that can be taken was presented and the procedure of the experiment was outlined. It was found that a single investigator can make reliable judgements of the severity of stuttering. Three separate severity ratings on the same reading passage were found to be highly correlated. These severity ratings correlated highly with reading rate, and to a slightly less extent, but still significant, with frequency counts of the moments of stuttering. Ratings on conversational speech of stutterers, and their own ratings of severity were also considered. It was felt that the therapist should not attempt to equate her assessment of stuttering behaviour and any change towards improvement with those assessments of change that the stutterer himself might make.

### Opsomming

Die verslag handel oor die verskillende soorte metings wat van haggelgedrag gedoen kan word, en hul verhouding teenoor mekaar. 'n Oorsig van die verskillende metings wat gedoen kan word, word verstrekk, en die prosedure wat ten grondslag van die proefneming lê, word geskets. Daar is bevind dat 'n enkele navorsingswerker 'n betroubare oordeel oor die erns van haggel kan vel. Daar is bevind dat daar 'n hoë mate van korrelasie tussen drie afsonderlike ernstigheidsbeoordelings van dieselfde leesgedeelte bestaan. Daar was 'n hoë korrelasie tussen hierdie ernstigheidsbeoordelings en die lees-tempo, en, in 'n mindere maar nog steeds betekenisvolle mate, die frekwensietellings op die oomblikke dat daar gehaggel is. Beoordelings van konwersasie-haggelaars, en hul eie ernstigheidsbeoordeling is ook in ag geneem. Daar word gemeen dat die terapeut geen poging moet aanwend om haar eie evaluasie van haggelgedrag en enige verandering in die rigting van verbetering gelyk te stel met dié evaluasies van veranderings wat deur die haggelaar self gedoen word nie.

### REFERENCES

1. Aron, M. L. (1964): *The Effects of the Combination of Trifluoperazine and Amylobarbitone on Adult Stutterers*. Ph.D. Thesis, University of the Witwatersrand, Johannesburg.
2. Aron, M. L. (1965): *The Effects of the Combination of Trifluoperazine and Amylobarbitone on Adult Stutterers*. Medical Proceedings, **II**, 227.
3. Bloodstein, O. (1959): *A Handbook on Stuttering for Professional Workers*. National Society for Crippled Children and Adults, Inc., Chicago.
4. Boehmler, R. M. (1958): *Listener Responses to Non-fluencies*. J. Speech Hearing Research, **I**, 132.

*Tydskrif van die Suid-Afrikaanse Logopediese Vereniging, Vol. 14, Nr. 1: Sept. 1967*

5. Cullinan, W. L., Prather, E. M., and Williams, D. E. (1963): *Comparison of Procedures for Scaling Severity of Stuttering*. J. Speech Hearing Research, **6**, 187.
6. Fairbanks, G. (1940): *Voice and Articulation Drillbook*. New York: Harper and Brothers.
7. Johnson, W. (1961): *Measurements of Oral Reading and Speaking Rate and Disfluency of Adult Male and Female Stutterers and Non-Stutterers*. J. Speech Hearing Disorders, Monograph Supplement **7**, 1.
8. Johnson, W., Darley, F. L., and Spriestersbach, D. C. (1963): *Diagnostic Methods in Speech Pathology*. New York: Harper and Row.
9. Johnson, W., and Knott, J. R. (1937): *The Distribution of Moments of Stuttering in Successive Readings of the Same Material*. J. Speech Disorders, **2**, 17.
10. Newman, P. W. (1963): *Adaptation Performances of Individual Stutterers: Implications for Research*. J. Speech Hearing Research, **3**, 223.
11. Quarrington, B. (1956): *Cyclical Variation in Stuttering Frequency and Some Related Form of Variation*. Canadian J. Psychology, **10**, 179.
12. Sander, E. K. (1961): *Reliability of the Iowa Speech Disfluency Test*. J. Speech Hearing Disorders, Monograph Supplement **7**, 21.
13. Sherman, D. (1952): *Clinical and Experimental Use of the Iowa Scale of Severity of Stuttering*. J. Speech Hearing Disorders, **17**, 316.
14. Sherman, D. (1955): *Reliability and Utility of Individual Ratings of Severity of Audible Characteristics of Stuttering*. J. Speech Hearing Disorders, **20**, 11.
15. Sherman, D., and Lewis, D. (1951): *Measuring the Severity of Stuttering*. J. Speech Hearing Disorders, **16**, 320.
16. Sherman, D., and Trotter, W. D. (1956): *Correlation Between Two Measures of the Severity of Stuttering*. J. Speech Hearing Disorders, **21**, 426.
17. Sherman, D., Young, M., and Gough, K. (1958): *Comparison of Three Measures of the Severity of Stuttering*. Proceedings of the Iowa Academy of Science, **65**, 381.
18. Thurstone, L. L. (1959): *The Measurement of Values*. Chicago: The University of Chicago Press.
19. Van Riper, C. (1962): Personal Communication.
20. Van Riper, C., and Hull, C. J. (1934): *The Quantitative Measurement of the Effect of Certain Situations on Stuttering*. Chapter 8, in *Stuttering in Children and Adults*. Ed., Johnson, W., Minneapolis: University of Minnesota Press, 1955.
21. Williams, D. E., and Kent, L. R. (1958): *Listeners Evaluations of Speech Interruptions*. J. Speech Hearing Research, **1**, 124.
22. Williams, D. E., Wark, M., and Minifie, F. D. (1963): *Ratings of Stuttering by Audio, Visual, and Audiovisual Cues*. J. Speech and Hearing Research, **6**, 91.
23. Young, M. A. (1961): *Predicting Ratings of Severity of Stuttering*. J. Speech Hearing Disorders, Monograph Supplement **7**, 31.