Journal of Speech and Hearing Disorders, Volume 49, 53–58, February 1984

STUTTERING AND SPEECH NATURALNESS

RICHARD R. MARTIN SAMUEL K. HAROLDSON KATHLEEN A. TRIDEN
University of Minnesota, Minneapolis

The present study was concerned with the development and evaluation of a scale of speech naturalness. Speech samples were recorded of the typical speech of 10 stutterers, 10 stutterers speaking without stuttering under 250-ms delayed auditory feedback (DAF), and 10 nonstutterers speaking normally. Using a 9-point scale, 30 unsophisticated listeners judged how natural the speech sounded in each sample. Results indicated that the stutterer samples were judged as sounding significantly more unnatural than the nonstutterer samples, and the DAF stutter-free samples were judged as sounding significantly more unnatural than the nonstutterer samples. The stutterer and DAF stutter-free samples were not judged as sounding significantly different in terms of speech naturalness. Interrater reliability, interrater agreement, and rater consistency for judging speech naturalness were all satisfactory.

Many current treatment programs for stuttering utilize feedback modification, prolongation, rhythm, or rate control in an effort to help stutterers reduce or eliminate instances of stuttering. Frequently, however, these procedures effect changes in the overall speech pattern resulting in a situation where "successfully" treated stutterers are relatively fluent, but their speech sounds slow, paced, or monotonous and can be discriminated from the speech of nonstutterers. To the extent that speech naturalness, as perceived by both the stutterer and his or her listeners, is an important consideration in recent treatment programs for stuttering, this aspect of speech needs additional experimental attention.

Several procedures have been used to study the naturalness of stutterers' fluent speech following treatment. In one method, observers were asked to judge whether a speech sample sounded natural or unnatural, or normal or abnormal (Ingham & Packman, 1978; Jones & Azrin, 1969). In another procedure, observers were requested to differentiate nonstuttered speech samples of stutterers from those of nonstutterers (Ingham & Packman, 1978; Runyan & Adams, 1978, 1979). In a third procedure, observers were asked to rate or identify aspects of stutterers' nonstuttered speech that were assumed to reflect speech naturalness or normalcy—rate, fluency, or prosody (Ingham & Packman, 1978; Perkins, Rudas, Johnson, Michael, & Curlee, 1974). Mallard and Meyer (1979) obtained information about the speech of treated and untreated stutterers by requesting observers to rate speech samples on a scale of most preferred speech pattern, second most preferred speech pattern, and least preferred speech pattern.

In the research cited above, observers were not required to scale and quantify their perceptions of speech naturalness. However, if speech naturalness is to be used meaningfully in studies dealing with various clinical and experimental treatments for stuttering, it must be determined empirically whether speech naturalness is a useful and scalable phenomenon. Scaling naturalness should provide a means for differentiating, in terms of numerical scale values, between both groups and individuals. The procedure should provide for differentiation, in terms of numerical scale values, among various stages or phases of treatment. Finally, if scaling speech naturalness is to be a useful procedure, observers must be able to scale the perception consistently and reliably. The present experiment was an initial effort to obtain information about scaling the speech naturalness of stutterers. Specifically, speech samples from nonstuttering speakers, stutterers, and stutterers speaking under delayed auditory feedback with no perceptible stutterings were rated by unsophisticated listeners on the dimension of speech naturalness.

METHOD

Speech Samples

Speech samples of approximately 1 minute in duration were obtained from 10 nonstutterers, 10 stutterers, and 10 stutterers speaking under 250-ms delayed auditory feedback (DAF). The nonstutterer group consisted of 6 men and 4 women aged 21–45. The stutterer group contained 6 men and 4 women aged 20–53. The group of stutterers who spoke under DAF contained 5 men and 5 women aged 20–51. The stutterers who spoke under DAF were not currently, nor had they been in the past, exposed to a long-term DAF therapy program. Of the 10 DAF-group stutterers, 5 had participated previously in a study which involved a brief exposure to DAF. The remaining 5 DAF-group subjects had no experience with DAF. The purpose of the present study was not to evaluate the effectiveness of DAF as a clinical treatment for stuttering. Rather, DAF was used as a means to obtain stutter-free speech samples in a relatively short period of time.

All speakers were seated alone in an experimental room and spoke spontaneously about any topic for 10

© 1984, American Speech-Language-Hearing Association

0022-4677/84/4901-0053$01.00/0

Downloaded From: http://jshd.pubs.asha.org/ by a Northern Illinois University User on 09/07/2016
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

minutes. At the end of 10 minutes, speakers under DAF were presented, via earphones, approximately 250-ms continuous delayed auditory feedback at what each speaker judged to be a comfortable loudness level. These speakers spoke under DAF for an additional 10 minutes. The speech of all speakers was recorded on an Ampex 440 Audiotape recorder located in an adjoining control room. A short 1000-Hz tone was recorded on the tape at 1-minute intervals.

A 1-minute segment was selected at random from each of the 10 tape recordings of the nonstutterers. The only restriction on the random selection was that a 1-minute segment not contain excessive pause or nontalking time. A 1-minute segment was also selected from the 10-minute audio recording of each stutterer. The selection was made so that, in the judgment of the experimenter, the 10 segments represented a range of stuttering frequencies. From the tape recording of each DAF speaker, a 1-minute segment was identified which, in the judgment of the experimenter, contained no instances of stuttering. The 30 1-minute speech segments (10 stutterer, 10 nonstutterer, 10 DAF) were randomized and dubbed onto a master rating tape with a 7-second silence between segments. Three orders of the rating tape were prepared. The original tape was the first order. The second order was prepared by placing the first 10 samples of the original tape at the end of the tape. A third order was obtained by placing the first 10 samples of the second tape at the end of the tape. The same three additional practice samples were placed at the beginning of each tape. Raters were not told the first three samples were practice samples, and the practice samples were numbered consecutively with the rating samples. After the three rating tapes, including the practice samples, were prepared, the sample number was recorded just before each sample.

Verbatim transcripts were prepared of each 1-minute speech sample, and six graduate students in communication disorders independently listened to the tape recording and marked each stuttered word on the transcript. None of the observers was familiar with the experiment. Observers were not provided with a definition of stuttering but were asked simply to underline any word they considered to be stuttered. Two observers each identified one instance of stuttering in the speech samples of the 10 nonstutterers. Three observers each identified one instance and one observer identified two instances of stuttering in the speech samples of the DAF speakers. These identifications were sufficiently infrequent to warrant the conclusion that the speech samples of the nonstutterers and DAF speakers were essentially free of perceptible stutterings. For the stutterer samples, the mean percentages of words marked as stuttered by the six observers for the 10 samples were 45.7, 45.2, 35.5, 30.9, 28.8, 19.0, 16.9, 7.8, 4.1, and 3.4. These results suggest that the experimenter was successful in selecting speech samples from stutterers that represented a wide range of stuttering frequency. It is the case, however, that the average percentage of words stuttered for the group was relatively high.

Raters and Rating Environment

Thirty undergraduate students recruited from beginning public speaking classes served as raters. Ten raters were assigned randomly to each rating tape order. Raters were seated in an experimental room connected to an adjoining control room via a one-way mirror and intercom system. In the experimental room, four armchairs were arranged in a semicircle around a loudspeaker (Electro-voice Marquis). All chairs were the same distance from the loudspeaker. The loudspeaker was connected to an Ampex 440 audiotape recorder in the control room.

Procedure

From one to four raters participated in a single session. Raters were given packets that contained a typed sheet of instructions followed by 11 sheets, each of which contained three rating scales. Each rating scale was a horizontal line approximately 6 inches long (15.24 cm). Vertical lines were placed at each end of the horizontal line, and seven additional vertical lines were spaced evenly between the ends. The vertical lines were numbered 1–9. Above the "1" was typed "highly natural," and above the "9" was typed "highly unnatural." The speech sample number was typed above each rating scale, and these numbers corresponded to sample numbers recorded on the rating tape.

The instruction sheet read as follows:

> We are studying what makes speech sound natural or unnatural. You will hear a number of 1-minute speech samples. The samples will be separated by a few seconds of silence. Each sample will be introduced by the sample number. Your task is to rate the naturalness of each speech sample. If the speech sample sounds highly natural to you, circle the 1 on the scale. If the sample sounds highly unnatural, circle the 9 on the scale. If the sample sounds somewhere between highly natural and highly unnatural, circle the appropriate number on the scale. Do not hesitate to use the ends of the scale (1 or 9) when appropriate. "Naturalness" will not be defined for you. Make your rating based on how natural or unnatural the speech sounds to you.

After the instructions were read, the experimenter entered the control room and activated the audiotape recorder. The 33 speech samples (3 practice, 30 rating tape) were played without stopping.

Rerate

After a rater completed the initial rating session, that person was scheduled for a second session to be conducted at least 1 week later. No information was provided about the nature of the task involved in the second session. In the rerate sessions the rating tape, equipment, instructions, environment, and procedures employed with each rater were the same as those used in the initial session. All raters accomplished the rerate task between 1 and 3 weeks following the initial rating.

Downloaded From: http://jshd.pubs.asha.org/ by a Northern Illinois University User on 09/07/2016
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

RESULTS

Naturalness Ratings

The mean, range, and standard deviation of naturalness scale values assigned by the 30 raters to the 10 stutterer, 10 nonstutterer, and 10 DAF speech samples and the frequency with which each scale value was assigned in the stutterer, nonstutterer, and DAF samples are given in Table 1. The data in Table 1 were submitted to an analysis of variance (Winer, 1971) in which Condition (stutterer, nonstutterer, DAF) was a "within" factor and Order (tape order 1, 2, 3) was a "between" factor. The main effect for Condition was significant ($F = 748.05$; $df = 2, 54$; $p < .01$), the main effect for Order was significant ($F = 4.16$; $df = 2, 27$; $p < .01$), and the Condition by Order interaction was significant at the .05 level ($F = 3.40$; $df = 4, 54$). A Tukey (Kirk, 1968) analysis was performed to test the significance of the difference in mean naturalness ratings between Orders 1 and 2, Orders 2 and 3, and Orders 1 and 3, for the stutterer, nonstutterer, and DAF samples. None of the F-values obtained from the nine Tukey tests reached a p value less than .08. The most conservative interpretation of these statistical results is that the relatively small observed differences in mean scale values among the tape orders probably are best explained as random variability rather than as a reliable difference among the three tape orders.

Newman-Keuls (Winer, 1971) tests were performed to test the significance of the observed differences in mean naturalness scale values between the stutterer and nonstutterer samples, the stutterer and DAF samples, and the nonstutterer and DAF samples in each of the tape orders. The results of these tests indicated that for each of the three tape orders, the mean naturalness rating assigned to the stutterer samples was significantly ($p < .01$) higher (more unnatural) than the mean rating assigned the nonstutterer samples. The mean rating assigned the DAF samples was significantly ($p < .01$) higher than the mean rating for the nonstutterer samples. The difference between mean ratings assigned the stutterer and DAF samples was not statistically significant ($p > .05$). Results of the analyses described above indicate that raters judged the speech samples of stutterers as significantly more unnatural than the speech samples of nonstutterers. The raters also judged the speech samples of stutterers under 250-ms DAF as significantly more unnatural than the speech samples of nonstutterers. The raters did not judge the speech samples of stutterers speaking under DAF as significantly different, in terms of speech naturalness, from the speech samples of stutterers.

Word Output and Naturalness Ratings

The mean numbers of words spoken in the approximately 1-minute samples of stutterer, DAF, and nonstutterer speakers were 66.1, 70.7, and 116.6, respectively. Pearson correlation coefficients were calculated between the number of words spoken in a speech sample and the mean naturalness rating assigned to that sample. The resulting correlations were stutterer, −.83; nonstutterer, −.51; DAF, −.59. The correlation for stutterer (−.83) was significant at the .01 level; the other two (−.51 and −.59) were not significant at the .05 level.

Stuttering Frequency and Naturalness Ratings

As indicated previously, the percentage of words stuttered in the stutterer speech samples ranged from 3.4 to 45.7. A correlation coefficient was calculated between the percentage of words stuttered in a given stutterer's speech sample and the mean naturalness rating assigned to that sample. The resultant correlation was .81. This correlation was significant beyond the .01 level.

Rater Reliability

In this study, interrater reliability and interrater agreement were defined in a manner suggested by Tinsley and Weiss (1975). Interrater reliability represents the extent to which naturalness ratings of the various speech samples assigned by different observers were proportional when expressed as deviations from the mean rating of the group. Although interrater reliability indicates how well judges agreed that the speech samples had the same relationship to each other in terms of naturalness scale values, this measure of reliability does not necessarily indicate to what extent observers agreed in terms of absolute scale values. Interrater agreement, on the other hand, indicates the extent to which all raters assigned the same naturalness scale value to a given speech sample.

In the present study, interrater reliability was determined by the intraclass correlation (R) procedure (Winer,

TABLE 1. Mean, range, standard deviation, and total naturalness scale values assigned by 30 raters to 10 stutterer, 10 nonstutterer, and 10 DAF speech samples.

Speech samples	$\bar{x}$	*Range*	*SD*	*Naturalness scale value* 1	2	3	4	5	6	7	8	9
Stutterer	6.52	1–9	2.00	3	8	17	26	33	41	66	46	60
250-ms DAF	5.84	2–9	1.79	0	15	17	42	42	67	58	45	14
Nonstutterer	2.12	1–5	1.17	123	78	48	42	9	0	0	0	0

Downloaded From: http://jshd.pubs.asha.org/ by a Northern Illinois University User on 09/07/2016
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

TABLE 2. Intraclass correlations for the mean speech naturalness ratings of 30 raters (R_{30}) and the average individual rater (R_1) for the stutterer, nonstutterer, and DAF speech samples.

Speech sample	R_{30}	R_1
Stutterer	.98*	.74**
Nonstutterer	.75**	.10
DAF	.98*	.57

*Significant at the .01 level.
**Significant at the .05 level.

1971). *R* utilizes an analysis-of-variance technique to determine the proportion of the total variance in a set of ratings due to the variance in the samples being rated. An *R* of 1.00 indicates very high interrater reliability; an *R* of 0.00 indicates the absence of interrater reliability. Table 2 gives the intraclass correlations for the mean reliability of the 30 raters (R_{30}) and for the average individual rater (R_1) on the stutterer, nonstutterer, and DAF speech samples. Correlations in Table 2 indicate that as a group the 30 raters were highly reliable in their naturalness ratings for the stutterer and DAF samples. The raters were somewhat less reliable in their naturalness ratings of the nonstutterer samples. This probably was due in large part to the restricted range of ratings assigned to the nonstutterer samples. In contrast to the high group reliability, the low intraclass correlations for the average individual rater (R_1) indicate that individual raters in the present study were unreliable in their ratings of speech naturalness. The one exception to this finding is the stutterer samples, for which the reliability of the average individual rater approached being satisfactory (R_1 = .74).

Rater Agreement

As previously stated, interrater agreement concerns the extent to which all raters assigned the same numerical naturalness scale value to a given speech sample. In order to determine interrater agreement in the present experiment, the naturalness scale value assigned by a given observer to a given speech sample was compared with the scale values assigned to that sample by every other observer. For the 30 raters who assigned naturalness ratings to the 10 stutterer speech samples, rater agreement was computed as follows: The scale value assigned by Rater 1 to the first speech sample was compared with the scale value assigned to that sample by each of the other 29 raters. Next, the scale value assigned by Rater 2 to the first speech sample was compared with the scale value assigned to that sample by each of Raters 3–30. Next, comparisons were made between Rater 3 and each of Raters 4–30. This procedure was continued until the scale value assigned by every rater to the first stutterer sample was compared with the scale value assigned to that sample by every other rater. The same procedure then was repeated for the other nine stutterer speech samples. This yielded a total of 4,350 comparisons of

TABLE 3. Cumulative number and percentage (in parentheses) of interrater agreements of speech naturalness scale values for stutterer, nonstutterer, and DAF samples. Total possible agreements for each speech sample group equals 4,350.

Speech sample	*Identical*	*Agreement (in scale values)* ±1	±2	±3	±4
Stutterer	1,386 (32)	3,223 (74)	4,023 (93)	4,304 (99)	4,350 (100)
Nonstutterer	1,362 (31)	3,286 (75)	4,109 (94)	4,350 (100)	4,350 (100)
DAF	1,394 (32)	3,337 (77)	4,199 (97)	4,350 (100)	4,350 (100)

naturalness scale values assigned by two different raters to the same sample for the 10 stutterer speech samples and 30 different listeners. Table 3 shows the extent of agreement for the 4,350 pairs of comparisons for the stutterer, nonstutterer, and DAF speech samples. The data in Table 3 indicate that in 1,386 (32%) of the 4,350 pairs of comparisons for the stutterer samples, the two raters assigned identical scale values. In 3,223 (74%) of the 4,350 comparisons, the two raters assigned values that were the same or that did not differ by more than plus or minus one scale value. The remaining entries in Table 3 are read in a similar manner. One meaningful summarization of the data in Table 3 is that in approximately 75% of the interrater comparisons for all three speech sample groups, raters assigned naturalness scale values that were the same or did not differ by more than one.

Rater Consistency

Rater consistency concerns the extent to which the rater assigns similar scale values to a given speech sample on two or more rating occasions. In the present study, rater consistency was expressed in terms of rate-rerate agreement. As indicated previously, all raters judged the same speech samples using the same procedures on two different occasions separated by at least 1 week. Table 4 gives the cumulative number and percentage of rate-rerate agreements for each speech sample group. For example, 30 raters assigned speech naturalness ratings to 10 stutterer samples (300 ratings) on each of two occasions. On 135 (45%) of the ratings, the raters assigned identical scale values on the two rating occasions. On 270 (90%) of the ratings, the raters assigned scale values that were identical or did not differ by more than plus or minus one scale value on the two rating occasions. On 292 (97%) of the ratings, the raters assigned scale values that were identical or did not differ by more than plus or minus one scale value on the two rating occasions. On 292 (97%) of the ratings, the raters assigned scale values that were identical or did not differ by more than plus or minus two scale values on the two rating occasions. In general, raters in the present study were consistent in their ratings of speech naturalness. Raters did not differ by more than plus or minus one scale value on 85% or more of their rate and rerate occasions.

Downloaded From: http://jshd.pubs.asha.org/ by a Northern Illinois University User on 09/07/2016
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

TABLE 4. Cumulative number and percentage (in parentheses) of rate-rerate agreements of speech naturalness scale values for stutterer, nonstutterer, and DAF speech samples. Total possible agreements for each speech sample group equals 300 (10 samples by 30 raters).

Speech sample	*Agreement (in scale values)*				
	Identical	*±1*	*±2*	*±3*	*±4*
Stutterer	135	270	292	298	300
	(45)	(90)	(97)	(99)	(100)
Nonstutterer	161	266	292	296	300
	(54)	(89)	(97)	(99)	(100)
DAF	131	256	296	300	300
	(44)	(85)	(99)	(100)	(100)

DISCUSSION

In the present study, speech naturalness appeared to be a useful and scalable aspect of the speech of both stutterers and nonstutterers. It is not particularly surprising that samples of stuttered speech were judged by observers as sounding much more unnatural than samples of speech from nonstutterers. It is also well known that stutterers speaking under the influence of 250-ms DAF typically sound unusual, although stutter-free (Goldiamond, 1965). This is documented in this study in that speech samples from stutterers speaking under delayed auditory feedback sounded almost as unnatural to listeners as samples of stuttered speech, and this in spite of the fact that the speech samples of the stutterers under delayed auditory feedback were essentially devoid of perceptible instances of stuttering. It was not the purpose of this study to evaluate the effectiveness of DAF as a clinical treatment for stuttering. Rather, the most important aspect of the present experiment was the demonstration that listeners can reliably scale a dimension of speech called speech naturalness.

These results have significant implications for the study and treatment of stuttering. As indicated earlier, many current therapy programs employ prolonged speech, masking, DAF, or rhythm to achieve stutter-free speech. Frequently, these treatment programs incorporate procedures designed to shape, if necessary, the stutter-free speech to more acceptable or desirable levels of rate, prosody, normalcy, naturalness, and so forth. Yet, how normal or natural stutterers' speech sounds to listeners has not been studied systematically or extensively. This is in spite of the fact that the speech naturalness of treated stutterers is a frequently voiced concern in the stuttering therapy literature.

Perhaps one reason the role of naturalness or normalcy in the treatment process for stutterers has not been studied extensively or systematically is the lack of a convenient means for quantifying this aspect of speech. Results of the present study suggest that observers can attend to and quantify a dimension of speech labeled speech naturalness. If future research corroborates this finding, then speech naturalness can be studied systematically. Experiments can be designed to determine the role of speech naturalness in the carry-over process. Studies can be conducted to delineate the relationships of speech rate, loudness, pitch, and inflection to perceived speech naturalness. In addition, procedures can be developed to manipulate the perceived speech naturalness of stutterers whose posttreatment speech is relatively free of stuttering. The relationship between speech rate (words spoken) and speech naturalness in the present study provides a good example of the potential usefulness of scaling speech naturalness. As indicated by the relatively high and negative correlation coefficient for the stutterer sample, speech naturalness ratings were influenced by the number of words spoken in the sample. In both the stutterer and DAF samples, the mean number of words spoken was rather small, and presumably this low word output was one reason the observers rated both kinds of speech samples as sounding quite unnatural. In the case of the stutterer samples, however, the naturalness ratings were also highly correlated with percentage of words stuttered. The DAF samples, however, were devoid of perceptible stutterings, yet these samples were rated as sounding almost as unnatural as the stutterer samples. A reliable procedure for scaling speech naturalness directly is potentially very useful in this situation. Procedures could be employed to increase the word rate of the DAF speakers, but at the same time retain the stuttering frequencies at zero. If the resultant speech is rated by listeners as relatively natural, then presumably the effect of DAF which caused the speech to sound unnatural, albeit stutter-free, was to slow speech rate. If, on the other hand, the resultant DAF speech at an increased word rate continues to be perceived as relatively unnatural, then it would seem that some characteristics other than, or in addition to, rate are implicated in the low naturalness ratings of the DAF speakers, and appropriate remedial procedures could be devised.

In the present study, raters were accurate, reliable, and consistent in their ratings of speech naturalness. Individual observers agreed with themselves reasonably well on two separate occasions. The degree to which raters agreed among themselves in terms of the absolute naturalness scale value assigned to a speech sample was relatively high for scaling experiments of this type. As a group, raters were highly reliable in their naturalness ratings, especially for the stutterer and DAF speech samples. Reliability of the individual raters, however, was quite low, except for the marginally satisfactory reliability of the speech naturalness ratings assigned by the average individual rater to the stutterer samples. The low reliability for individual raters observed in the present study has implications for future experiments involving the scaling of speech naturalness. In group parametric studies in which judgments of speech naturalness are obtained and in which observer accuracy is assessed and reported in terms of interrater reliability, it probably would be prudent to utilize groups of observers. In single-subject experiments and clinical case studies, however, it may not be necessary to employ large numbers of observers. In single-subject experiments and case

Downloaded From: http://jshd.pubs.asha.org/ by a Northern Illinois University User on 09/07/2016
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx

studies the concern typically is not so much with how observers proportionally order a group of speakers in terms of speech naturalness (interrater reliability) as it is with accuracy of the absolute naturalness scale values assigned to a given speaker across situations and over time. The high interrater agreement and rater consistency data obtained in the present study suggest that it is possible for individual observers to scale speech naturalness in an accurate and consistent fashion. Additional research is needed, however, to determine whether a single observer can accurately and consistently scale the speech naturalness of a given stutterer over time and across situations.

ACKNOWLEDGMENTS

This study was supported in part by the Bryng Bryngelson Communication Disorders Research Fund, University of Minnesota. The authors wish to thank Marilee Williams and Dorothy Martin for their assistance with the experiment.

REFERENCES

GOLDIAMOND, I. (1965). Stuttering and fluency as manipulatable operant classes. In L. Krasner & L. Ullman (Eds.), *Research in behavior modification*. New York: Holt, Rinehart, & Winston.

INGHAM, R. J., & PACKMAN, A. C. (1978). Perceptual assessment of normalcy of speech following stuttering therapy. *Journal of Speech and Hearing Research, 21*, 63–73.

JONES, R. J., & AZRIN, N. H. (1969). Behavioral engineering: Stuttering as a function of stimulus duration during speech synchronization. *Journal of Applied Behavior Analysis, 2*, 223–229.

KIRK, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Wadsworth.

MALLARD, A. R., & MEYER, L. (1979). A Listener preference for stuttered and syllable-timed speech production. *Journal of Fluency Disorders, 4*, 117–121.

PERKINS, W., RUDAS, J., JOHNSON, L., MICHAEL, W. B., & CURLEE, R. F. (1974). Replacement of stuttering with normal speech. III Clinical effectiveness. *Journal of Speech and Hearing Disorders, 39*, 416–428.

RUNYAN, C. M., & ADAMS, M. R. (1978). Perceptual study of the speech of "successfully therapeuticized" stutterers. *Journal of Fluency Disorders, 3*, 24–39.

RUNYAN, C. M., & ADAMS, M. R. (1979). Unsophisticated judges' perceptual evaluation of the speech of "successfully therapeuticized" stutterers. *Journal of Fluency Disorders, 4*, 29–38.

TINSLEY, H. E. A., & WEISS, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology, 22*, 353–376.

WINER, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Received September 13, 1982
Accepted August 17, 1983

Request for reprints should be sent to Richard Martin, Ph.D., Department of Communication Disorders, 115 Shevlin Hall, 164 Pillsbury Drive SE, University of Minnesota, Minneapolis, MN 55455.

© 1984, American Speech-Language-Hearing Association

0022-4677/84/4901-0058$01.00/0

Downloaded From: http://jshd.pubs.asha.org/ by a Northern Illinois University User on 09/07/2016
Terms of Use: http://pubs.asha.org/ss/rights_and_permissions.aspx