Is Listener Comfort a Viable Construct in Stuttering Research?

RESEARCH NOTE

Sue O'Brian
Ann Packman
Mark Onslow
Angela Cream
Nigel O'Brian
University of Sydney, Sydney, New South Wales, Australia

Kaely Bastock
University of Queensland, Brisbane, Queensland, Australia

This article reports the development of a tool for measuring how comfortable a person feels when communicating with someone who has undergone treatment for stuttering. The person rates the speaker on a 9-point Listener Comfort Scale (9 = *extremely comfortable*, 1 = *extremely uncomfortable*). In a preliminary investigation of the reliability and validity of the scale, 15 unsophisticated listeners rated video recordings of 10 adults before and after a prolonged-speech treatment for stuttering and of 10 matched controls. The results were compared with those of another 15 listeners who rated the same recordings with the widely used 9-point Speech Naturalness Scale (R. R. Martin, S. K. Haroldson, & K. A. Triden, 1984). Results showed that reliability of the Speech Naturalness Scale was superior to the Listener Comfort Scale, although users of both scales were able to distinguish between pretreatment speech, posttreatment speech, and the speech of controls. The results suggest that the Listener Comfort Scale captures information that is somewhat different than the information captured by the Speech Naturalness Scale. The authors concluded that the concept of listener comfort is a potentially useful additional way of investigating the social validity of behavioral treatments for stuttering.

KEY WORDS: stuttering, social validity, listener comfort, outcome, speech naturalness

The best known and most effective methods for behavioral control of chronic stuttering are based on the use of the novel speech pattern known as prolonged speech (PS) and its variants (for reviews, see Ingham, 1984; Onslow, 1996; Packman, Onslow, & Menzies, 2000). Traditionally, objective measures of stuttering and speech rate have been used to indicate the success of such treatments. However, it has long been recognized that such behavioral measures alone do not provide sufficient information about the outcomes of these treatments.

Consequently, those who have researched methods for the behavioral control of stuttering have sought to supplement objective behavioral speech measures with measures of social validity. Such measures attempt to quantify the value of the behavioral change that has been achieved with treatment. Social validation studies involve either comparisons of clients' posttreatment speech with that of nonstuttering control speakers or subjective judgments of changes in speech through the pretreatment to posttreatment period. Together, behavioral and social validity measures provide more extensive evidence of the successful outcome of behavioral treatments for stuttering.

The early social validation studies of PS treatments showed that listeners are consistently able to distinguish the speech of individuals who have participated in such treatments from that of normally fluent

speakers (Ingham & Packman, 1978; Perkins, Rudas, Johnson, Michael, & Curlee, 1974; Runyan & Adams, 1978, 1979). In an attempt to quantify the differences that might be responsible for the distinctiveness of the posttreatment speech of those who stutter, Martin, Haroldson, and Triden (1984) developed a 9-point scale to measure speech naturalness. Subsequent studies showed this naturalness scale to be valid and reliable (Ingham, Gow, & Costello, 1985; Ingham, Ingham, Onslow, & Finn, 1989; Onslow & Ingham, 1987). Social validation studies have since used this scale to evaluate the naturalness of posttreatment speech of adults who stutter in relation to the speech of normally fluent speakers (Ingham et al., 1985; Metz, Schiavetti, & Sacco, 1990; Onslow, Hayes, Hutchins, & Newman, 1992; Runyan, Bell, & Prosek, 1990) and in relation to pretreatment speech of adults who stutter (Franken, Boves, Peters, & Webster, 1992; Kalinowski, Noble, Armson, & Stuart, 1994).

Measures of speech naturalness focus on how one dimension of speech is perceived by the listener. However, it is arguable that a more important outcome of treatment, at least to the client, is the extent to which treatment increases the social acceptability of speech. One of the first studies to address this issue—albeit with a novel speech pattern other than PS—found that unsophisticated listeners preferred the sound of stuttered speech to the sound of syllable-timed speech of some clients (Mallard & Meyer, 1979).

Another early study (Cullinan, Prather, & Williams, 1963) assessed the concept of easiness-to-listen-to, however, this was only in relation to pretreatment stuttering severity. The authors' interest was not to develop a tool to measure the social validity of treatment outcomes but to compare different methods of evaluating pretreatment stuttering severity. However the authors did note that a 7-point scale using this concept produced reliable ratings.

Franken, Van Bezooijen, and Boves (1997) proposed the notion of the communicative suitability of posttreatment speech. In their study, listeners rated on a 10-point scale the suitability of clients' speech for 10 various speaking contexts outside the clinic. Franken et al. found that the stuttering participants attracted lower communicative suitability ratings than did the control speakers, but the stuttering participants' speech was deemed more acceptable after treatment than before treatment in all communication contexts.

Finally, Susca and Healey (2001, 2002) addressed the concept of how comfortable people feel listening to stuttered and nonstuttered speech. In a laboratory study, they digitally manipulated a sample of stuttered speech to produce a number of samples with varying amounts of stuttering, including one sample with all stuttering removed. Ten listeners heard the samples, along with a sample from a normally fluent speaker, and made judgments on a 7-point Likert scale about listener comfort. This scale discriminated between the samples, with listeners feeling more comfortable listening to the "treated sample" than to any of the stuttered samples or to the normally fluent sample. Although Susca and Healey's study was a laboratory study, using the digitally manipulated speech of one speaker only, it suggests that the concept of listener comfort may have considerable clinical potential. It captures the sense of listeners' feelings of what it would be like to communicate with a speaker.

The purpose of this study was to investigate the viability of adapting Susca and Healey's (2001, 2002) Listener Comfort Scale for use as a stuttering outcome measure in research and clinical practice. This measure could potentially address clinicians' and clients' concerns that posttreatment nonstuttered speech, especially if unnatural sounding, may in fact be less acceptable to the general population than the pretreatment stuttered speech. It was considered essential to establish that the concept of listener comfort is independent, at least to some extent, of the concept of speech naturalness. If a Listener Comfort Scale provides information that is no different from the information provided by the naturalness scale, then its contribution to outcome evaluation would be minimal.

In this study, listener comfort was rated on a 9-point scale so that it could be compared with the 9-point naturalness scale. In the Listener Comfort Scale, however, a high score indicates a high degree of listener comfort, in contrast to the naturalness scale on which a high score indicates a low degree of speech naturalness.

Method

Listeners

Listeners in this study were 30 members of the nonacademic staff of the University of Sydney in Sydney, Australia. Each was randomly assigned to one of two groups of 15 listeners. None had occupations where they met or interacted with people who had communication disorders. Potential participants who were acquainted with a person with a speech, language, or hearing disorder were excluded. There were 12 men and 18 women listeners, with approximately equal numbers assigned to each group. Their occupations included administrative assistant, receptionist, telephonist, gardener, and carpenter.

Speakers

The speakers used in this study were 10 adults who stutter and 10 normally fluent controls matched for age

(within 2 years) and gender. The stuttering participants were the first 10 clients to reach the maintenance stage in a trial of a variant of a PS treatment program known as the Camperdown Program, in which PS is learned without programmed instruction (O'Brian, Cream, Onslow, & Packman, 2001; O'Brian, Onslow, Cream, & Packman, 2002). Seven of the speakers were men and 3 were women. Speakers ranged in age from 18 to 51 years. Pretreatment stuttering rate taken from the study samples ranged from 1.6 percent stuttered syllables (%SS) to 46%SS, and the posttreatment rate ranged from 0%SS to 1.6%SS. The 10 control speakers were chosen from the administrative and maintenance staff of a major Sydney teaching hospital. All 10 control speakers were screened for communication problems, and English was the first language for each of them. One of the clients had Spanish as a first language but was a fluent speaker of English.

Stimulus Tape

Two videotapes were constructed, each of which contained the same 43 recordings, each 30 s in length, presented in a different random order. The 43 recordings consisted of (a) a pretreatment and a posttreatment sample of each of the 10 clients, (b) two samples from each of the 10 controls, and (c) three practice samples at the beginning of each video.

Each of the client samples consisted of the first 30 s of a 10-min, within-clinic video recording, which was free from interruption or comment by the clinician and which was considered by the treating clinician to be representative of the overall conversation. The samples had been digitally recorded with arc lighting and a lapel microphone. The pretreatment recordings were made approximately 1 month before the beginning of treatment. The posttreatment recordings were made approximately 3 months after the clients' entry into the maintenance stage of the program. Any reference to speech or treatment in the recording was excluded. Two different 30-s samples were chosen from each control recording, with the speaker talking about a different topic in each case. Two samples from each control were used so that there would be equal numbers of samples from both groups; however, only the first sample for each control was used in the data analysis. The three practice samples were of a client pretreatment, a client posttreatment, and a control speaker.

Procedure

Listeners were randomly allocated to either Part A or Part B of the study.

Part A

The 15 listeners watched and scored the samples for listener comfort (LC) using a 9-point, equal-interval scale (9 = *extremely comfortable*, 1 = *extremely uncomfortable*). Half of the listeners watched one randomly ordered version of the videos and the remainder watched the other randomly ordered version. Listeners watched the tape independently on a video monitor in a quiet room with an investigator present. Before beginning the task, participants were given a score sheet and written and verbal instructions. The instructions for the Listener Comfort Scale (see the Appendix) were loosely based on those used by Martin et al. (1984).

Part B

The 15 other listeners watched the same two videos, but rated them for speech naturalness (NAT) using the 9-point Martin et al. (1984) Speech Naturalness Scale. These listeners followed the same procedure as in Part A except that the instructions used were the same as those in the Martin et al. study.

Results

Descriptive Statistics for LC and NAT Ratings

Tables 1 and 2 show the mean, range, and standard deviation of the Listener Comfort Scale and Speech Naturalness Scale scores assigned to each sample. They also show the %SS pretreatment and posttreatment for each speaker. In most cases, the standard deviation and range of scores for the Listener Comfort Scale is higher than for the Speech Naturalness Scale.

A Pearson correlation coefficient was calculated for the pretreatment Listener Comfort Scale and the Speech Naturalness Scale ratings, the posttreatment ratings, and the controls' ratings. These were –.96, –.49, and –.86, respectively, and show that although the pretreatment correlation between the Listener Comfort Scale and the Speech Naturalness Scale was extremely high, it was only moderate for posttreatment scores and was low with the outlying data point removed. This would indicate that, after treatment at least, the two scales are measuring something different. A negative correlation coefficient in this case indicates a positive correlation between the two scales, as the values for the two scales were reversed.

Interrater Reliability and Agreement

Intraclass correlations (ICC) (2,1) (Shrout & Fleiss, 1979) were calculated for the Listener Comfort Scale and the Speech Naturalness Scale scores for the three population samples (pretreatment, posttreatment, and control) combined. The ICC value for the Listener Comfort Scale was .5 and for the Speech Naturalness Scale was .71. Although both scores are in the fair-to-good

Table 1. Mean (*M*), range, and standard deviation (*SD*) of the Listener Comfort Scale scores for each speaker, and mean pretreatment and posttreatment %SS scores.

	Pretreatment				Posttreatment				Controls		
Speaker	*M*	Range	*SD*	%SS	*M*	Range	*SD*	%SS	*M*	Range	*SD*
1	6.00	8	2.14	3.7	8.07	5	1.22	0.0	8.33	5	1.05
2	5.07	7	1.75	12.3	5.73	8	2.22	1.0	8.13	4	0.99
3	5.87	7	1.88	5.0	8.20	4	1.01	0.0	8.53	3	0.64
4	3.20	6	1.70	46.0	7.67	4	0.98	0.8	8.53	3	0.83
5	6.20	6	1.86	1.6	6.73	7	1.71	0.0	7.33	6	1.50
6	5.00	7	2.07	7.7	7.80	4	1.15	0.0	7.80	4	1.08
7	5.73	7	2.05	9.5	6.93	7	1.75	1.6	7.80	6	1.32
8	4.80	9	2.18	11.4	7.33	6	1.68	0.0	8.67	2	0.49
9	3.60	6	2.06	27.6	8.33	5	1.11	0.0	8.07	3	0.80
10	4.33	8	2.16	20.0	7.27	5	1.58	1.5	8.20	4	1.08
M	4.98	7.1	1.99	14.5	7.41	5.5	1.44	0.5	8.14	4	0.98

Table 2. Mean (*M*), range, and standard deviation (*SD*) of the Speech Naturalness Scale scores for each speaker, and mean pretreatment and posttreatment %SS scores.

	Pretreatment				Posttreatment				Controls		
Speaker	*M*	Range	*SD*	%SS	*M*	Range	*SD*	%SS	*M*	Range	*SD*
1	4.53	7	1.92	3.7	1.53	3	0.74	0.0	1.40	4	0.91
2	5.40	6	1.72	12.3	3.07	8	1.79	1.0	1.33	4	0.82
3	5.20	6	1.57	5.0	1.27	3	0.59	0.0	1.27	2	0.46
4	8.20	4	1.08	46.0	2.73	5	1.39	0.8	1.00	1	0.00
5	3.53	8	1.77	1.6	3.27	6	1.49	0.0	2.07	4	0.96
6	5.53	6	1.55	7.7	2.4	5	1.30	0.0	1.80	4	0.77
7	4.33	7	2.13	9.5	2.53	6	1.25	1.6	1.33	3	0.62
8	5.33	7	1.88	11.4	1.47	2	0.52	0.0	1.07	2	0.26
9	7.27	5	1.33	27.6	2.87	6	1.64	0.0	1.47	3	0.74
10	6.33	6	1.59	20.0	1.73	6	1.44	1.5	1.27	4	0.80
M	5.57	6.2	1.65	14.5	2.29	5	1.22	0.5	1.4	3.1	0.63

range of .4–.75 specified by Fleiss (1986), it is obvious that the Listener Comfort Scale was less reliable than the Speech Naturalness Scale. Rater agreement was measured using Martin et al.'s (1984) procedure to determine the extent to which all listeners assign the same scale value to a given sample, by comparing the value assigned by each listener to the value assigned by every other listener. For the Listener Comfort Scale, 56% of ratings were assigned to within ±1 scale value, and for the Speech Naturalness Scale, 68% of ratings were assigned to within ±1 scale value. Here again, the Speech Naturalness Scale performed better.

Analysis of Variance (ANOVA) for the Pretreatment and Posttreatment Groups

A major purpose of this study was to evaluate the Listener Comfort Scale for use as a tool to measure treatment outcome. Therefore, an ANOVA was completed on the pretreatment and posttreatment scores combined, without those of the control group. The results are shown in Table 3. Although the samples (speakers) themselves accounted for a large amount of the variance (42.28%), a fairly high variance (36.26%) could also be attributed to the raters alone. Reliability

Table 3. ANOVA results and estimated variance components for pretreatment and posttreatment scores on the Listener Comfort scale.

Source of variation	Sum of squares	Mean square	Estimated variance component	% total variance
Sample (s)	667.59	35.14	2.2660	42.28
Rater (r)	560.29	40.02	1.9435	36.26
Sample × Rater (sr)	306.91	1.15	1.1500	21.46

Table 4. ANOVA results and estimated variance components for pretreatment and posttreatment scores on the NAT scale.

Source of variation	Sum of squares	Mean square	Estimated variance component	% total variance
Sample (s)	1,184.70	62.35	4.0633	64.69
Rater (r)	248.59	17.76	0.8180	13.02
Sample × Rater (sr)	371.15	1.40	1.4000	22.29

was obviously higher for control samples than for client samples so with the control group removed the interjudge reliability coefficient reduced still further to .42.

A similar analysis was done for the NAT data without the controls (see Table 4). This shows that the samples (speakers) accounted for most of the variance (64.69%) while in this case the raters only accounted for 13.02% of the variance. This indicates significantly higher consistency among raters and, hence, a more reliable scale. The interjudge reliability coefficient was .65.

An intrajudge reliability coefficient was also calculated for these two groups of listeners using the method described by Cullinan et al. (1963). For the Listener Comfort Scale, the intrajudge reliability coefficient was .79 and for the Speech Naturalness Scale .78.

Pretreatment and Posttreatment Listener Comfort Scale and Speech Naturalness Scale Scores (Treatment Effect)

Table 1 shows that, for all clients, the mean Listener Comfort Scale rating assigned to posttreatment speech was higher than the mean rating for pretreatment speech, showing that the group of listeners generally indicated a higher level of listener comfort for posttreatment samples. Data for the group were statistically significant, $t(9) = -5.39$, $p < .001$. For 8 of the 10 clients, the mean posttreatment score was within one scale score of the mean for their matched control. Clients who evidenced some residual posttreatment stuttering tended to receive lower Listener Comfort Scale ratings. Similar trends were evident for Speech Naturalness Scale scores, with the data being statistically significant, $t(9) = 6.81$, $p < .001$. Speech Naturalness Scale scores for all clients improved posttreatment although only 5 of the 10 clients had achieved values within one scale score of their matched control.

Posttreatment and Control Listener Comfort Scale and Speech Naturalness Scale Scores

Both the Listener Comfort Scale and Speech Naturalness Scale scores distinguished between posttreatment and control samples. Both data sets were statistically significant: Listener Comfort Scale, $t(18) = -2.59$, $p = .018$; Speech Naturalness Scale, $t(18) = 3.52$, $p = .002$.

In terms of treatment outcome, the results for the treatment reported in this study (O'Brian et al., 2001, in press) differ markedly from the treatment results reported by Franken et al. (1992) and Kalinowski et al. (1994). In contrast to those reports, all clients in this study were found to show improved speech naturalness posttreatment, whereas clients from the Kalinowski et al. and Franken et al. studies were perceived as having either the same or worse speech naturalness following treatment. Although the perceived improvement was greater for individuals with more severe stuttering, the same trend was evident for both mild and severe stuttering.

Discussion

Existing outcome measures for PS treatments focus on speaker behaviors (stuttering rate and speech rate) and on a perceived dimension of speech naturalness. Consequently, there is a need to develop a social validation measure for outcome research that is based on how people experience clients' speech after treatment. For that purpose, the present study evaluated a scale of LC, similar to that used by Susca and Healey (2001).

Results demonstrated, at best, only moderate reliability for use of the LC scale with stuttering clients. This result is surprising when compared with the study by Cullinan et al. (1963). When they assessed the easiness-to-listen-to of 20-s samples as rated by unsophisticated listeners on a 7-point scale, their interjudge reliability coefficient was relatively high (.72) compared to that obtained in this study (.42). This difference cannot be attributed to differences in intrajudge reliability for the two studies because the reliability was in fact the same for both studies (.79). The major differences between the Cullinan et al. study and the present one are the use of videotaped samples and a 9-point scale in the present study. Other studies have shown little difference between 7- and 9-point scales; hence, it is possible that the absence of visual images of particpants had a beneficial impact on reliability. We would argue, however, that video recordings are more socially valid than audio-only recordings. Alternatively, it may have meant that the easiness-to-listen-to concept meant something quite different to participants than the comfortable-to-listen-to concept.

In terms of ICC and pairwise agreement scores, the group reliability for the Listener Comfort Scale was inferior to the Speech Naturalness Scale. Again, the intrajudge reliability for both scales was similar (.79, .78). These results are not surprising, however, considering that the Speech Naturalness Scale elicits perceptual judgments of how speech sounds, whereas the new Listener

Comfort Scale elicits a value judgment about how comfortable a listener feels with the client's speech. Presumably, more variables contribute to judgments of listener comfort than to judgments of speech naturalness. Susca and Healey (2002) concluded from their study that "listeners appear to bring certain idiosyncratic listener comfort requirements to a listening situation" (p. 154). Indeed, Table 1 shows that, in comparison to the Speech Naturalness Scale, the Listener Comfort Scale consistently elicited a wider range of scores.

However, as an indicator of the impact of treatment on the listener, the LC scale may still prove to be useful. For a group of listeners, it was useful in differentiating between clients' pretreatment and posttreatment speech, and between clients' posttreatment speech and the speech of control participants. As an indicator of treatment effect, this information is more useful than obtaining absolute scores on a scale. These results seem robust, for several reasons. First, for every client, listeners felt more comfortable listening to posttreatment speech than to pretreatment speech. Second, the mean Listener Comfort Scale pretreatment–posttreatment difference of 2.43 was statistically and, apparently, clinically significant. Pairwise agreement data show that the majority of listeners disagreed by two or fewer scale values, so this effect cannot be accounted for by random variation in Listener Comfort Scale scoring among the listeners. Finally, each of these results was replicated with the Speech Naturalness Scale, which is known to be valid.

The fact that the Listener Comfort Scale results were consistent the Speech Naturalness Scale raises the issue that the two scales may be measuring the same speech dimension. However, there are reasons to suggest that this may not be the case. In the first instance, the listeners for the Listener Comfort Scale and the Speech Naturalness Scale scoring were given completely different instructions about their listening task: The instructions for the Listener Comfort Scale group were designed to focus attention on their experience of comfort and the instructions for the Speech Naturalness Scale group were designed to focus their attention on the naturalness of the speakers. Second, the correlation between the Listener Comfort Scale and the Speech Naturalness Scale was high for pretreatment samples but only moderate for posttreatment samples. These results suggest that the Listener Comfort Scale and the Speech Naturalness Scale are not measuring exactly the same speech dimension.

In conclusion, the results demonstrate support for the concept of listener comfort in studies of outcomes of behavioral treatments for stuttering, and they suggest a potential role for the development of a new speech construct in such studies. This LC concept seems to be valid in capturing, to some extent, a dimension other than speech naturalness. It is also important because it taps into information of prime importance to the client; whether the average speech partner is more comfortable when listening to pretreatment or posttreatment speech, and when listening to posttreatment or normally fluent speech. This is a potentially valuable addition to outcome research because it quantifies whether the stuttering treatment has led to a positive change in terms of how the client's speech affects the listener. This may be considered a truly valid measure of social outcome.

Further research is needed, however, to determine how best to use and evaluate listener comfort. It could be that if listeners are asked to make a forced choice, using the method of Ingham and Packman (1978), more reliable data could be attained. In other words, judges would be confronted with a pretreatment and a posttreatment recording of the client's speech, or a posttreatment recording and a matched control, and asked to choose the sample that they felt more comfortable listening to. This procedure could even be incorporated clinically, at the end of treatment, to demonstrate to the client that conversational partners are more comfortable with posttreatment speech. It could also be useful in comparing outcomes of different treatments by asking listeners to identify the treatment outcomes with which they are most comfortable.

Acknowledgments

This study was supported by the vacation studentship scheme of the University of Sydney. The authors would also like to thank the staff members who gave their time to participate in this project.

References

Cullinan, W., Prather, E., & Williams, D. (1963). Comparison of procedures for scaling severity of stuttering. *Journal of Speech and Hearing Research, 6,* 187–194.

Fleiss, J. (1986). *The design and analysis of clinical experiments*. New York: Wiley.

Franken, M., Van Bezooijen, R., & Boves, L. (1997). Stuttering and communicative suitability of speech. *Journal of Speech, Language, and Hearing Research, 40,* 83–94.

Franken, M. C., Boves, L., Peters, H. F. M., & Webster, R. L. (1992). Perceptual evaluation of the speech before and after fluency shaping stuttering therapy. *Journal of Fluency Disorders, 17,* 223–241.

Ingham, R. J. (1984). *Stuttering and behavior therapy: Current status and experimental foundations.* San Diego, CA: College-Hill Press.

Ingham, R. J., Gow, M., & Costello, J. M. (1985). Stuttering and speech naturalness: Some additional data. *Journal of Speech and Hearing Disorders, 50,* 217–219.

Ingham, R. J., Ingham, J. C., Onslow, M., & Finn, P. (1989). Stutterers' self-ratings of speech naturalness: Assessing effects and reliability. *Journal of Speech and Hearing Research, 32,* 419–431.

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.

Kalinowski J., Noble, S., Armson, J., & Stuart, A. (1994). Pretreatment and posttreatment speech naturalness ratings of adults with mild and severe stuttering. *American Journal of Speech-Language Pathology, 3*, 61–66.

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

Martin, R. R., Haroldson, S. K., & Triden, K. A. (1984). Stuttering and speech naturalness. *Journal of Speech and Hearing Disorders, 49,* 53–58.

Metz, D. E., Schiavetti, N., & Sacco, P. R. (1990). Acoustic and psychophysical dimensions of the perceived speech naturalness of nonstutterers and posttreatment stutterers. *Journal of Speech and Hearing Disorders, 55,* 516–525.

O'Brian, S., Cream, A., Onslow, M., & Packman, A. (2001). A replicable, nonprogrammed, instrument-free method for the control of stuttering with prolonged-speech. *Asia Pacific Journal of Speech, Language, and Hearing, 6,* 91–96.

O'Brian, S., Onslow, M., Cream, A., & Packman, A. (in press). The Camperdown Program: Speech outcomes for a short, operationalized, nonprogrammed, instrument-free method for control of chronic stuttering. *Journal of Speech, Language, and Hearing Research.*

Onslow, M. (1996). *Behavioral management of stuttering.* San Diego, CA: Singular.

Onslow, M., Hayes, B., Hutchins, L., & Newman, D. (1992). Speech naturalness and prolonged-speech treatments for stuttering: Further variables and data. *Journal of Speech and Hearing Research, 35,* 274–282.

Onslow, M., & Ingham, R. J. (1987). Speech quality measurement and the management of stuttering. *Journal of Speech and Hearing Disorders, 52,* 2–17.

Packman, A., Onslow, M., & Menzies, R. (2000). Novel speech patterns and the control of stuttering. *Disability and Rehabilitation, 22,* 65–79.

Perkins, W. H., 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 treated stutterers. *Journal of Fluency Disorders, 3,* 25–40.

Runyan, C. M., & Adams, M. R. (1979). Unsophisticated judges' perceptual evaluations of the speech of "successfully treated" stutterers. *Journal of Fluency Disorders, 4,* 29–38.

Runyan, C. M., Bell, J. N., & Prosek, R. A. (1990). Speech naturalness ratings of treated stutterers. *Journal of Speech and Hearing Disorders, 55,* 434–438.

Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Susca, M., & Healey, E. C. (2001). Perceptions of simulated stuttering and fluency. *Journal of Speech, Language, and Hearing Research, 44,* 61–72.

Susca, M., & Healey, E. C. (2002). Listener perceptions along a fluency-disfluency continuum: A phenomenological analysis. *Journal of Fluency Disorders, 27,* 135–161.

Received June 25, 2002

Accepted November 12, 2002

DOI: 10.1044/1092-4388(2003/041)

Contact author: Sue O'Brian, Australian Stuttering Research Centre, Faculty of Health Sciences, University of Sydney, East Street (Gate 2), Lidcombe, New South Wales, Australia 2141. E-mail: s.obrian@fhs.usyd.edu.au

Appendix. Instructions for Part A of the study.

We are studying how comfortable people feel listening to the speech of others. The speech of some people may make you feel quite comfortable, while the speech of others may make you feel quite uncomfortable.

On this video, there are 43 thirty-second samples of various people speaking. The first three are practice samples. After watching the whole of each sample, you are to indicate on the Record Form how comfortable you would feel listening to the person's speech in a social situation. Your response should reflect your feelings about the way the person was speaking (i.e., how comfortable you would feel listening to them), not what the person was saying or how their personality affected you.

If you felt extremely comfortable listening to the person's speech, then circle 9 on the scale. If you felt extremely uncomfortable, then circle 1. If you felt somewhere between extremely comfortable and extremely uncomfortable, circle the appropriate number between 9 and 1 on the scale. Do not hesitate to use any number on the scale but do not record in the spaces between the numbers. In other words, make all your judgments correspond to a number from 1 to 9. There are no right or wrong answers, only a personal judgment. You should circle any number that you think matches how you felt.

"Comfortable" will not be defined for you. Just circle the number that best indicates how comfortable you felt about the person's speech, in that sample.