Self-Rating of Stuttering Severity as a Clinical Tool

Sue O'Brian
Ann Packman
Mark Onslow
Australian Stuttering Research Centre,
The University of Sydney, Australia

Scaling is a convenient and equipment-free means for speech-language pathologists (SLPs) and clients to evaluate stuttering severity in everyday situations. This study investigated the extent to which the severity ratings of 10 adult stuttering speakers, made immediately after speaking and again from recordings 6 months later, agreed with ratings made by an SLP. For 9 of the 10 speakers, there was good agreement between their initial ratings and those of the SLP. For 8 of the 10 speakers, there was also good agreement between their initial ratings and those made from recordings 6 months later, indicating that the severity ratings made at the time of speaking were reliable. These findings support the use of the 9-point scale as a clinical measurement procedure.

Key Words: stuttering, measurement, severity, scales, self-reporting

Considerable research has been done on the use of rating scales by both clinicians and unsophisticated listeners to measure stuttering severity (see Ingham, 1984). Early studies by Sherman and colleagues (Lewis and Sherman, 1951; Sherman, 1952, 1955; Sherman & Trotter, 1956) evaluated a 9-point scale and found this to be a reliable tool for assessing stuttering severity. Later studies highlighted the fact that whether a scale had 5, 7, 9, or 15 points seemed to make little difference to mean scale values or reliability (Cullinan, Prather, & Williams, 1963; Curran & Hood, 1977). Further, there is little difference in scores when points on the scale are clearly defined (Cullinan et al., 1963), when participants are given repeated exposure to the task or feedback about group mean scores (Young, 1969a, 1969b), or when live as opposed to recorded samples are used (Cullinan et al., 1963). Finally Lewis and Sherman (1951), Young (1961), and Martin (1965) concluded that background experience of the rater (unsophisticated listeners compared with experienced clinicians) was not a significant variable, although this result was not subsequently confirmed by Onslow, Andrews, and Costa (1990) or by Eve, Onslow, Andrews, and Adams (1995). In both the latter studies, experienced clinicians had higher interjudge and intrajudge agreement than that of unsophisticated listeners or generalist clinicians. In summary, severity rating scales are a valid and reliable method for evaluating stuttering severity by both experienced and inexperienced listeners.

There is also a large body of literature on the use of self-evaluation for quantifying stuttering and related speech behaviors. For example, studies have reported on the self-evaluation of number of stutters and rate of speech during the treatment process (Harrison, Onslow, Andrews, Packman, & Webber, 1998; Ingham, 1982; Onslow, Costa, Andrews, Harrison, & Packman, 1996) and the benefit of such procedures in reducing the amount of stuttering, regardless of the accuracy of these measures (Costello, 1975; LaCroix, 1972). In this context, self-evaluation appears to assist in achieving and maintaining treatment benefits. In a laboratory experiment, Ingham and Cordes (1997a) also investigated self-judgments of stuttering behavior in different situations and compared these results with those of experienced observers. Poor intrajudge reliability for many of the judges complicated the interpretation of their findings. The positive effects of self-judged expectations of success on actual performance have also been reported (Hillis & McHugh, 1998; Ornstein & Manning, 1985). Further studies have shown that adults who stutter can reliably rate both the sound and feel of their own speech (Finn & Ingham, 1994; Ingham, Ingham, Onslow, & Finn, 1989) using a 9-point speech naturalness scale (Martin, Haroldson, & Triden, 1984). Both these studies have shown reasonable agreement between participants' ratings made while speaking and ratings made from recordings and those of experienced clinicians. Therefore, there is already some evidence that

self-measurement techniques are a useful means for reporting speech performance.

Ingham and Cordes (1997b) reviewed the pertinent literature on self-measurement and suggested that "self-judgments should play a central role in any clinically valid measurement procedure for stuttering" (p. 414). Given the reliability of severity rating scales when used by clinicians and unsophisticated listeners and given that they are easy and convenient (equipment free) to use, possibly with little or no training, it is surprising that little research has been done on the use of these scales by people who stutter themselves. There have been only two studies of self-rating of stuttering severity, and these were published over 30 years ago (Aron, 1967; Naylor, 1953).

In the Naylor study (1953), 24 adults who stuttered were trained to use a 9-point severity rating scale, ranging from 1 *(least severe stuttering)* to 9 *(most severe stutter-ing)*. The participants then rated the severity of their own stuttering while reading aloud for 63 s. This reading task was audiotaped, and participants subsequently rated the severity of a 9-s segment of their own recording. Their ratings were compared with those of 13 graduate students who had also been trained to use the scale. Results showed a correlation of .76 between participants' self-ratings while reading and those of the students. A correlation of .62 was achieved between participants' self-ratings made later from recordings and those of the students.

In the Aron study (1967), 46 adults who stuttered read aloud the Rainbow Passage (Fairbanks, 1960) and rated the severity of their stuttering immediately afterward on a 9-point scale, ranging from 1 (*no stuttering*) to 9 (*very severe stuttering*). The experimenter rated severity on the same scale. Thirty minutes later, participants listened to an audiotape of their own reading and rerated severity. Correlation between participants' ratings made immediately after reading and those from recordings was .50. Correlations between the experimenter's ratings and the participants' first and second ratings were not as high as in Naylor (1953; i.e., .66 and .52, respectively). The participants were not trained in this study.

The results of these two studies, however, along with the literature on the use of severity rating scales generally, suggest that such scales may be used by adults who stutter to rate the severity of their own stuttering during the treatment process. This would be a useful clinical tool to supplement traditional stutter-count measures, particularly for reporting on stuttering severity outside the clinic in the person's everyday speaking environment. These ratings could inform the treatment process by providing valuable information about the generalization of treatment effects in many different situations. Typically such information has been collected through the use of audiotape. However, audiotaping is intrusive, so it is not always possible, or even feasible, for clients to record themselves speaking in many situations, and hence, *selected situations* may not truly reflect overall performance.

This study therefore is an investigation of the self-administration of a 9-point severity rating scale. The primary aim was to see how well self-ratings of stuttering severity, made across a range of everyday speaking situations by adults who stutter, would agree with ratings made by a treating clinician. Good agreement between the two sets of scores would indicate that client ratings could be used in the clinical setting without training, whereas poor agreement would indicate that training might be required. It was considered important in addressing this issue to look at individual client ratings because of possible individual differences: the ratings of some clients might agree with those of the speech-language pathologist (SLP), but those of others might not agree. It was also considered important to investigate the use of severity ratings across a range of speaking contexts, both within and outside the clinic, as would occur in clinical use of the scale.

Two supplementary investigations were conducted. First, the clients rated their stuttering severity from audiotapes 6 weeks later, to see how well these ratings agreed with the ratings they made immediately after completing the speech task. Second, the study provided an opportunity to investigate differences in stuttering across different speaking situations. Thus, measures of stuttering rate, namely percentage syllables stuttered (%SS), were also made from the recordings.

Specifically, the study used a 9-point scale to answer the following questions: (a) What percentage of client ratings made immediately after completing a conversation (*immediate* ratings) would agree within 1 point with SLP ratings? (b) What percentage of client immediate ratings would agree within 1 point with their ratings made from audiotapes 6 months later (*tape* ratings)? (c) Do severity ratings and stuttering frequency measures detect similar trends in stuttering?

Method

Participants

Participants were 10 stuttering adults (speakers) on a waiting list for treatment and one SLP. None of the speakers had any experience using a stuttering rating scale, although 5 had participated in treatment programs that involved identification of stuttering events. Because this involved comparisons with clinician judgments, it could be interpreted as training and hence as altering their own perception of severity. There were 9 men and 1 woman, ranging in age from 19 to 52 years.

The first author was the SLP in this study because she was the intended clinician for these stuttering clients for the clinical trial they were about to enter. She was experienced in the measurement and treatment of stuttering, having worked in specialist stuttering centers for more than 15 years. She was also familiar with severity rating scales, having used both 9- and 10-point scales for the clinical measurement of stuttering severity. To confirm that this SLP's responses to the task were typical of an experienced clinician, her scores were compared with those of a second judge (see *Reliability* section). This second judge was chosen because it was known from O'Brian, Packman, Onslow, and O'Brian (in press) that her scores using the same 9-point rating scale were both reliable (intrajudge reliability of .99) and representative of a group of 12

experienced SLPs (scores consistently in the mid-range for all clinicians).

Procedure

Part 1. The 10 speakers each made 5-min audio-recordings of themselves speaking in six different situations. Three of these recordings were made within the clinic, using a portable tape recorder (Sony Walkman Model WM-D6C) with lapel microphone: reading a 250-word passage, talking with a clinician, and talking on the phone to a stranger. The other three recordings were made outside the clinic in everyday speaking situations using a portable tape recorder (Sony Model TCM-453V): talking with a family member, talking to a friend, and talking on the phone to a person of choice.

During Part 1, participants were given a score sheet and the following written instructions:

Immediately after making each of the six recordings outlined on the accompanying sheet, you are asked to make a judgment about the severity of your stuttering in that situation. When making this judgment, you should base it on the following scale: 1 = NO STUTTERING; 9 = EXTREMELY SEVERE STUTTERING. If you felt there was no stuttering in your speech during that situation, then you would write "1" in the space provided. If you felt your speech would be classified as containing extremely severe stuttering, then you would write "9" in the space provided. If you felt that the severity of your stuttering was somewhere between "no stuttering" and "extremely severe stuttering," then you would write the appropriate number in the space provided. Do not hesitate to use any number from the scale, but do not use more than one number at a time. In other words, make all your judgments correspond to a number from 1–9. There are no right or wrong answers, only a personal judgment.

The severity scale was represented at the top of the score sheet by a horizontal line with nine marks placed at equal intervals along the line, numbered from 1 to 9. The ends were defined as above. Speakers then simply selected a number from 1 to 9 for each of the six recordings.

Six to 8 weeks after completing the task, an investigator recontacted the speakers unexpectedly and asked them to come in to the clinic again. During this visit, they listened to their own six recordings, in the order in which they were originally made, without headphones, using a desktop audiotape recorder. The previous instructions were modified slightly to reflect that they were now to listen to the recordings of their speech made earlier but were otherwise identical. They were instructed to assign a number to each recording in the same manner as before, using the same 9-point severity scale. On neither occasion were speakers trained in the use of the scale.

Part 2. The SLP rated the severity of stuttering on the 60 recordings from Part 2 (10 speakers × 6 situations) using the same scale. The SLP listened to each sample without headphones using the desktop audiotape recorder

that was used for the tape ratings by the speakers. The SLP had no knowledge of the speakers' self-ratings from Part 1.

As a supplementary analysis, this SLP also measured %SS for each sample using a button-press counter–timer device. The %SS task was completed after the severity rating task, so that measuring stuttering rate would not influence the ratings of severity.

Reliability. Approximately 6 months later, the SLP rerated the severity of all the recordings to assess intrajudge agreement. Table 1 shows that 83% of the SLP's pairs of ratings (50 of 60) agreed to within 1 point on the scale. To assess the interjudge reliability of this SLP's scores, a second SLP (second judge) also rated the severity of the 60 samples using the same 9-point scale in the same manner. Table 1 shows that 77% of the pairs of ratings for the two SLPs (46 of 60) agreed to within 1 point. The maximum difference between any two ratings was 3 points.

To assess interjudge agreement for the supplementary %SS scores, 10% of samples were also randomly chosen for repeat rating by the second judge. The correlation between the two sets of scores was .99. The maximum difference between any two ratings was 2.9%SS (15.9 – 13.0%SS) while the smallest difference was 0.1%SS (0.3 – 0.2%SS).

Results

The Appendix contains the raw scores for Parts 1 and 2.

Agreement Between Speakers' Immediate Ratings and SLP's Ratings

Figure 1 displays the amount and direction of agreement between SLP's and speakers' immediate ratings for each of the 10 speakers for each of the six speaking situations. Of speakers' immediate ratings, 78% (47 of 60) agreed with those of the SLP within 1 point. Of the 13 comparisons that differed by more than 1 point, 5 came from 1 speaker (S7). This speaker, who had had no previous therapy, consistently rated his stuttering as less severe than did the SLP (or the second judge). The remaining 8 comparisons that differed by more than 1 point were distributed randomly across the other speakers. With S7 removed from the analysis, 85% of comparisons were within 1 point. There was no clinically significant difference in agreement between the ratings of

Table 1. Intrajudge agreement for the SLP's first and second ratings and interjudge agreement for the SLP's and second judge's ratings.

Difference	Intrajudge agreement			Interjudge agreement		
	No.	%	Cumulative %	No.	%	Cumulative %
0	27	45	45	16	27	27
1	23	38	83	30	50	77
2	8	13	97	11	18	95
3	2	3	100	3	5	100

Note. Agreement is expressed as the difference in scale points. SLP = speech-language pathologist.

Figure 1. Severity scores for speakers' immediate ratings and initial speech-language pathologist's (SLP's) ratings for each of the six situations and 10 speakers. Ratings were made on a 9-point scale ranging from 1 *(no stuttering)* to 9 *(extremely severe stuttering)*. Vertical lines represent the difference between the two scores.
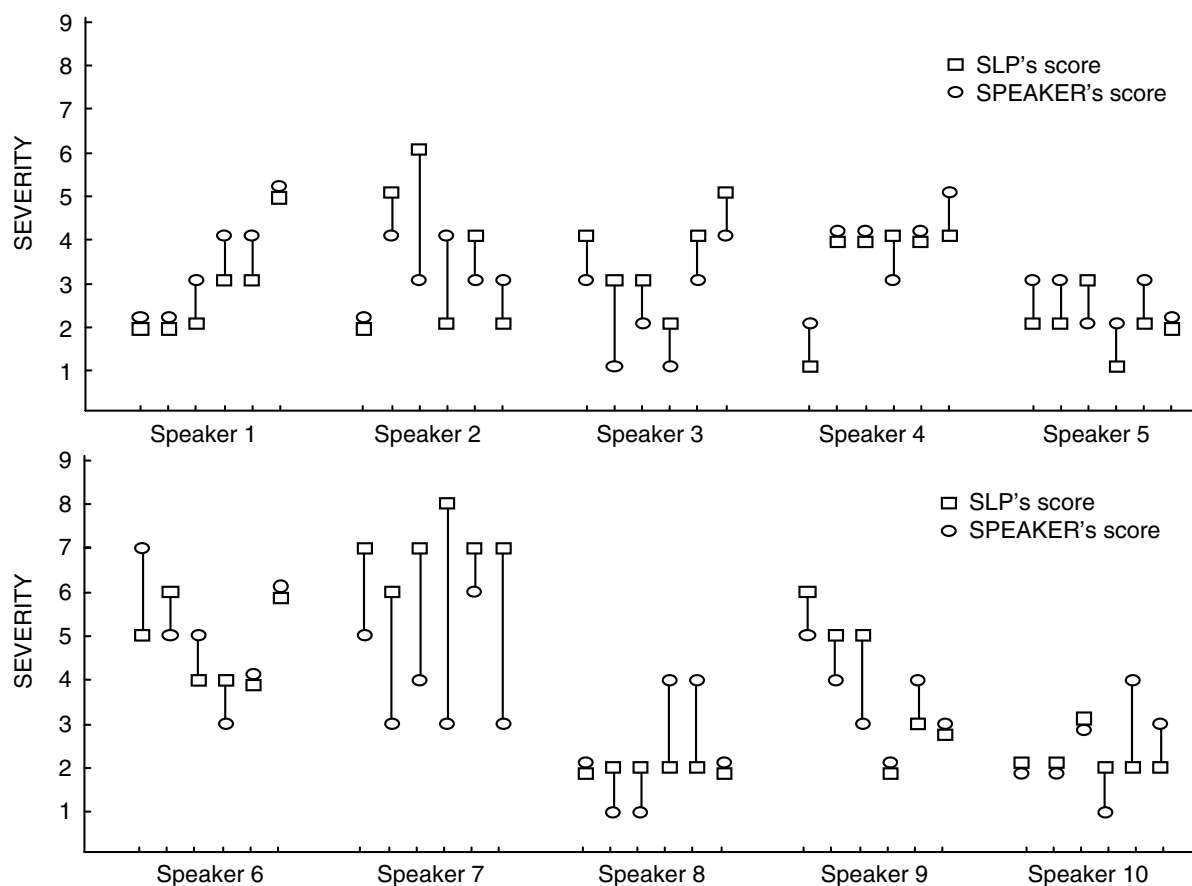
samples collected within the clinic and those of samples collected outside the clinic (23 and 24 of the comparisons, respectively, were within 1 point).

A further analysis was done between those speakers who had had previous treatment (S_T; $n = 5$) involving identification of stuttering events and those who had not (S_{NT}; $n = 5$). Twenty-six out of 30 S_T scores (86.6%) agreed to within 1 point of the SLP's scores, and 21 out of 30 S_{NT} scores (70%) agreed to the same extent. However, with S7 removed from the analysis, the scores were very similar ($S_T = 86.6\%$; $S_{NT} = 83.3\%$) for the two groups. In other words, for 9 out of 10 speakers (all but S7), there was good agreement between their immediate ratings and those of the SLP whether or not they had received previous treatment.

Agreement Between Speakers' Immediate Ratings and Tape Ratings

Table 2 shows that when speakers' ratings made immediately after speaking in each of the six situations were compared with the speakers' ratings made from the recordings of the same speech samples 6 weeks later, 75% (45 of the 60 comparisons) were within 1 point. Of the

remaining 15 comparisons, in most cases (13), the scores from tape ratings were higher (rated as more severe stuttering) than those of the immediate ratings, although 6 of these were from the same 2 speakers.

Agreement Between Speakers' Tape Ratings and SLP's Ratings

Given that speakers' tape ratings were consistently higher than their immediate ratings, a further analysis was conducted to see which of these agreed more closely with the SLP's ratings. Table 3 shows that agreement for tape ratings was lower (70% within 1 point) than that for immediate ratings (78% within 1 point), even though the SLP's ratings were all made under the same conditions as the speakers' tape ratings.

Stuttering Across Communication Contexts

Figure 2 shows the mean SLP's %SS scores and severity ratings for the 10 speakers, across the six communication contexts. The trends across situations, both within the clinic and outside the clinic, were very similar for both measures.

Table 2. Agreement between speakers' immediate ratings and tape ratings.

Difference	No.	%	Cumulative %
0	16	27	27
1	29	48	75
2	8	14	86
3	5	9	97
4	0	0	97
5	2	3	100

Note. Agreement is expressed as the difference in scale points.

Discussion

This study investigated the use of a 9-point scale by adults who stutter to rate the severity of their own stuttering. The primary aim was to establish the extent to which clients' ratings of stuttering severity, made both within and outside of the clinic, agreed with those of an SLP. Good agreement would indicate that the procedure could be useful for clients' reporting of stuttering severity in a wide range of everyday speaking situations.

With this in mind, the datum of initial interest was the percentage of pairs of comparisons that were within 1 point on the scale. Comparisons that differed by more than this could not be considered to be clinically useful for reporting beyond-clinic progress. Results indicated that the ratings of the SLP in this study agreed to a reasonable extent with clients' self-ratings of stuttering severity, whether or not the clients had received previous treatment involving identification of stuttering. Agreement (within 1 point on the scale) between the SLP's ratings and the speakers' immediate ratings was 78%. In practical terms, this would mean that approximately 8 out of every 10 judgments of stuttering severity between SLP and client could be expected to be comparable.

This result is perhaps not surprising when compared with Ingham and Cordes's (1997a) findings. Although the task in the present study was quite different, it involved making global judgments of severity as opposed to identifying intervals of speech as "stuttered" or "not stuttered," Ingham and Cordes found that there were many intervals which stuttering speakers and experienced judges agreed were "stuttered." This is not to say, however, that the two procedures will necessarily reflect similar levels of stuttering severity.

As might be expected, this finding of good agreement did not apply to all participants. One speaker in this study

Table 3. Agreement between SLP's ratings and speakers' tape ratings.

Difference	No.	%	Cumulative %
0	17	28	28
1	25	42	70
2	10	17	87
3	5	8	95
4	3	5	100

Note. Agreement is expressed as the difference in scale points.

Figure 2. Mean SLP's percentage syllables stuttered (%SS) scores (A) and severity ratings (B) for the 10 speakers within the clinic (WC) and outside the clinic (OC). The WC communicative contexts were reading aloud, talking with a clinician, and talking on the phone to a stranger; the OC contexts were talking with a family member, talking with a friend, and talking on the phone to a person of choice.
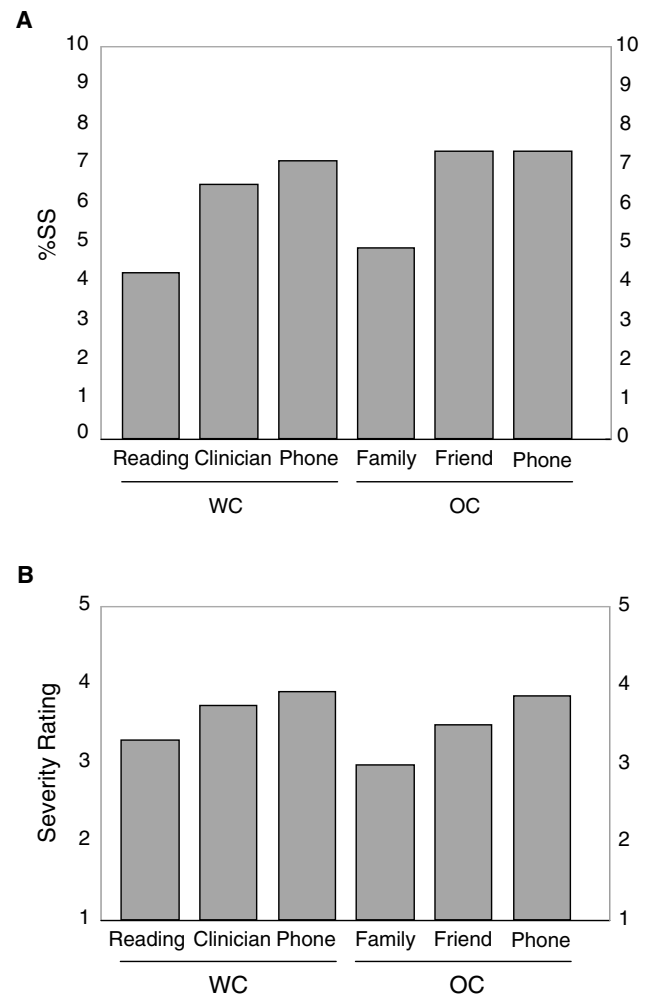
consistently rated his stuttering less severely than did the SLP and the second judge. Eleven of the 12 comparisons for the immediate ratings by this speaker (6 communication contexts × 2 SLPs, including reliability ratings from the second judge) differed by more than 1 point. In fact, this speaker's scores differed by up to 6 points on the scale from the SLPs' scores. In other words, for 9 of the 10 speakers, there was good agreement (85% of scores within 1 point) between SLPs' scores and those of the speakers' immediate ratings. However for the remaining speaker, there was very poor agreement (only 17%, or one rating, within 1 point). Ingham and Cordes (1997a) also noted that for some speakers there were consistent differences between their judgments of their own stuttering and those of an experienced observer.

For the 9 speakers with good agreement, 3 agreed with the SLP at an acceptable level of within 1 point for all six

communicative contexts. Four speakers differed from the SLP by more than 1 point, but for only one of the six communicative contexts each, and the last 2 speakers differed from the SLP by more than 1 point for only two of the six communicative contexts (see Figure 1).

This study also demonstrated that 75% of speakers' immediate/tape comparisons were within 1 scale point of each other, suggesting that for the majority of participants, judgments of stuttering severity made immediately after conversing were quite similar to those made when subsequently listening to a recording of their speech. This result included the speaker (S7) whose scores at the time of speaking were consistently lower than those of the SLP. For the 2 speakers for whom consistent discrepancies existed between scores for the two tasks, scores from tape ratings were always higher than scores from immediate ratings and less likely to correspond to those of the SLP. This observation was also noted by both Naylor (1953) and Aron (1967) in their respective studies. The fact that the majority of immediate/tape comparisons agreed quite well, however, adds support to the validity of immediate ratings.

Generally speaking, then, SLPs might reasonably expect that many clients' ratings, made at the time of speaking using the 9-point scale reported in this study, will agree to an acceptable level with their own ratings. However, it would seem advisable that SLPs establish the extent to which they agree with self-ratings for each client at the start of treatment, using methods based on those reported in this article. This can be done by having the client self-record speech outside the clinic, rate severity at the time of speaking, and discuss this rating with the clinician after they listen to the recording together.

There are several benefits to using self-ratings during treatment. First, if there is reasonable agreement between the client and the SLP, then they will be able to communicate easily and effectively about the severity of the client's stuttering in everyday communication contexts away from the clinic. This will provide the SLP with valuable information about generalization of treatment gains and whether there is a need for specific transfer activities. Second, clients will have a way of quantifying their stuttering severity as treatment progresses rather than relying solely on clinician judgment.

While this study shows that acceptable levels of agreement are likely, disagreement between client and SLP is not a basis for discarding the scale for that client. Indeed, disagreement can be used to advantage. Providing the disagreement is not due to erroneous use of the scale, it can form the basis for useful dialogue about factors that contribute to clinician's and client's judgments of severity. Ingham and Cordes (1997b) highlighted some of the issues relating to disagreement on self-reports of stuttering. It may be that the client's judgments are influenced by unobservable factors. On the one hand, the client's ratings may be inflated by the presence of word avoidance, or feelings of effort or anxiety, while on the other hand, they may be deflated if the client pays little attention to, or is not particularly concerned by, stuttering severity. It is possible that training in identification of stuttering events and/or in the self-use of the 9-point scale may improve

agreement between client and SLP. This would increase the tool's usefulness for reporting beyond-clinic severity. However, further experimental research is required to establish whether this is the case.

The fact that self-rating does not involve the use of intrusive equipment bypasses some of the potential problems associated with audiotaping, which necessarily limit the number and variety of contexts in which treatment data can be collected. Self-reporting of beyond-clinic stuttering severity with a rating scale means that speech can be evaluated in a wider and possibly more representative range of situations. Another problem with audiotape recorders is that they can become a discriminative stimulus. Discriminated learning occurs when stimuli associated with the treatment setting become stimuli for stutter-free speech. Rating machines and recording machines, for example, have been known to have this effect (Onslow, 1996). It is unknown whether a rating procedure may also act in the same way. Stuttering is known to be reactive to the act of measurement itself, and either method will certainly increase the client's awareness of the assessment procedure, but whether this is the critical factor in such situations needs further investigation.

This study also provided an opportunity to see if perceived stuttering severity varied across the six speaking situations. Interestingly, measures based on stutter counts (%SS) and severity ratings indicated similar trends across speaking situations, with no apparent difference in stuttering rate or severity for the groups between speech samples made in the clinic and those made in naturalistic situations. This finding suggests that SLPs should never assume that clients generally stutter either less or more in the clinic than they do outside the clinic, at least before treatment.

In summary, the 9-point severity scale investigated in this study provides a simple, convenient, and reasonably reliable tool for clients to provide information on their stuttering in naturalistic environments without the use of an audiotape recorder. This study has addressed the use of self-ratings as a measurement procedure that may be used during the therapy process. However, the finding that self-ratings of stuttering severity at the time of speaking can be made reasonably reliably for everyday speaking situations suggests that the 9-point scale has potential for measuring treatment outcome, at least as a supplementary measure. Self-report is currently considered to be a valuable supplement to the objective measures of stuttering rate used in outcome studies (see Boberg & Kully, 1994; Ingham & Cordes, 1997b; O'Brian, Onslow, Cream, & Packman, 2003; Packman, Onslow, O'Brian, & Huber, 2004). In the O'Brian et al. report, a self-report inventory that participants completed at the end of treatment (including participants' self-ratings of stuttering severity on the same 9-point scale as used in this study) indicated that they were stuttering at a higher rate than suggested by the objective measures made from audiotape recordings. The speech samples on which severity ratings were based in this study were collected before treatment and so contained considerable amounts of stuttering. Further research is required to establish whether the 9-point scale investigated in this study can be used with

similar levels of agreement after treatment, when clients' speech will be mostly stutter free.

Acknowledgments

References

Aron, M. L. (1967). The relationships between measurements of stuttering behaviour. *Journal of the South African Logopedic Society, 14,* 15–34.

Boberg, E., & Kully, D. (1994). Long-term results of an intensive treatment program for adults and adolescents who stutter. *Journal of Speech and Hearing Research, 37,* 1050–1059.

Costello, J. (1975). The establishment of fluency with time-out procedures: Three case studies. *Journal of Speech and Hearing Disorders, 40,* 216–231.

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

Curran, M. & Hood, S. (1977). Listener ratings of severity for specific disfluency types in children. *Journal of Fluency Disorders, 2,* 87–97.

Eve, C., Onslow, M., Andrews, C., & Adams, R. (1995). Clinical measurement of early stuttering severity: The reliability of a 10-point scale. *Australian Journal of Human Communication Disorders, 23,* 26–29.

Fairbanks, G. (1960). *Voice and articulation drillbook.* New York: Harper & Row.

Finn, P., & Ingham, R. J. (1994). Stutterers' self-ratings of how natural speech sounds and feels. *Journal of Speech and Hearing Research, 37,* 326–340.

Harrison, E., Onslow, M., Andrews, C., Packman, A., & Webber, M. (1998). Control of stuttering with prolonged speech: Development of a one-day instatement program. In A. Cordes, & R. J. Ingham. (Eds.), *Treatment efficacy in stuttering: A search for empirical bases* (pp. 191–212). San Diego, CA: Singular Publishing Group.

Hillis, J. W., & McHugh, J. (1998). Theoretical and pragmatic considerations for extraclinical generalization. In A. K. Cordes & R. J. Ingham (Eds.), *Treatment efficacy for stuttering: A search for empirical bases* (pp. 243–292). San Diego, CA: Singular Publishing Group.

Ingham, R. J. (1982). The effects of self-evaluation training on maintenance and generalization during stuttering treatment. *Journal of Speech and Hearing Disorders, 47,* 271–280.

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

Ingham, R., & Cordes, A. (1997a). Identifying the authoritative judgments of stuttering: Comparisons of self-judgments and observer judgments. *Journal of Speech, Language and Hearing Research, 40,* 581–594.

Ingham, R., & Cordes, A. (1997b). Self-measurement and evaluating stuttering treatment efficacy. In R. Curlee & G. Siegal (Eds.), *Nature and treatment of stuttering: New directions* (pp. 413–437). Boston, MA: Allyn & Bacon.

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.

LaCroix, Z. E. (1972). Management of disfluent speech through self-monitoring procedures. *Journal of Speech and Hearing Disorders, 38,* 272–274.

Lewis, D., & Sherman, D. (1951). Measuring the severity of stuttering. *Journal of Speech and Hearing Disorders, 16,* 320–326.

Martin, R. R. (1965). Direct magnitude-estimation judgments of stuttering severity using audible and audible-visible speech samples. *Speech Monographs, 32,* 169–177.

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

Naylor, R. V. (1953). A comparative study of methods of estimating the severity of stuttering. *Journal of Speech and Hearing Disorders, 18,* 30–37.

O'Brian, S., Onslow, M., Cream, A., & Packman, A. (2003). The Camperdown Program: Outcomes of a new prolonged-speech treatment model. *Journal of Speech, Language, and Hearing Research, 46,* 933–946.

O'Brian, S., Packman, A., Onslow, M., & O'Brian, N. (in press). Measurement of stuttering: A comparison of stuttering count and severity rating methods. *Journal of Speech, Language, and Hearing Research.*

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

Onslow, M., Andrews, C., & Costa, L. (1990). Parental severity scaling of early stuttered speech: Four case studies. *Australian Journal of Human Communication Disorders, 18,* 47–61.

Onslow, M., Costa, L., Andrews, C., Harrison, E., & Packman, A. (1996). Speech outcomes of a prolonged-speech treatment for stuttering. *Journal of Speech and Hearing Research, 39,* 734–749.

Ornstein, A. F., & Manning, W. H. (1985). Self-efficacy scaling by adult stutterers. *Journal of Communication Disorders, 18,* 313–320.

Packman, A., Onslow, M., O'Brian, S., & Huber, A. (2004). Measuring the outcomes of behavioral treatments: What are our data telling us? In A. K. Bothe (Ed.), *Evidence-based treatment of stuttering: Empirical issues and clinical implications.* Mahwah, NJ: Erlbaum.

Sherman, D. (1952). Clinical and experimental use of the Iowa scale of severity of stuttering. *Journal of Speech and Hearing Disorders, 17,* 316–320.

Sherman, D. (1955). Reliability and utility of individual ratings of severity of audible characteristics of stuttering. *Journal of Speech and Hearing Disorders, 20,* 11–16.

Sherman, D., & Trotter, W. D. (1956). Correlation between two measures of the severity of stuttering. *Journal of Speech and Hearing Disorders, 21,* 426–429.

Young, M. A. (1961). Predicting ratings of severity of stuttering. *Journal of Speech and Hearing Disorders, Monograph Supplement No. 7,* 31–54.

Young, M. A. (1969a) Observer agreement: Cumulative effects of rating many samples. *Journal of Speech and Hearing Research, 12,* 135–143.

Young, M. A. (1969b) Observer agreement: Cumulative effects of repeated ratings of the same samples and of knowledge of group results. *Journal of Speech and Hearing Research, 12,* 144–155.

Contact author: Sue O'Brian, PhD, Australian Stuttering Research Centre, The University of Sydney, P.O. Box 170, Lidcombe, New South Wales 1825, Australia. E-mail: S.Obrian@fhs.usyd.edu.au

Appendix

Stuttering Severity Ratings for the Six Communication Contexts

Speaker	Previous therapy	Location	Communication context	%SS	Speaker immed	Speaker tape	SLP 1st rating	SLP 2nd rating
1	Yes	WC	Reading	1.3	2	1	2	2
			Clinician	0.2	2	3	2	2
			Phone	0.6	3	3	2	2
		OC	Family	2.3	4	3	3	3
			Friend	0.7	4	4	3	2
			Phone	1.7	5	4	5	3
2	No	WC	Reading	0.5	2	3	2	2
			Clinician	8.3	4	6	5	6
			Phone	11.5	3	6	6	7
		OC	Family	1.8	4	3	2	3
			Friend	6.5	3	3	4	5
			Phone	5.0	3	2	2	3
3	No	WC	Reading	7.0	3	6	4	5
			Clinician	6.4	1	6	3	5
			Phone	7.7	2	7	3	4
		OC	Family	2.5	1	2	2	2
			Friend	6.5	3	4	4	4
			Phone	7.5	4	6	5	4
4	Yes	WC	Reading	0	2	2	1	1
			Clinician	6.3	4	4	4	5
			Phone	9.6	4	6	4	6
		OC	Family	7.7	3	6	4	4
			Friend	9.3	4	7	4	6
			Phone	9.6	5	8	4	5
5	No	WC	Reading	0.9	3	3	2	2
			Clinician	0.8	3	2	2	2
			Phone	0.4	2	2	3	2
		OC	Family	2.1	2	1	1	2
			Friend	1.3	3	2	2	2
			Phone	2.3	2	3	2	3
6	No	WC	Reading	6.4	7	7	5	6
			Clinician	6.3	5	6	6	6
			Phone	5.7	5	6	4	6
		OC	Family	3.9	3	4	4	4
			Friend	5.7	4	3	4	4
			Phone	9.6	6	6	6	6
7	No	WC	Reading	15.9	5	5	7	7
			Clinician	24.6	3	4	6	8
			Phone	17.7	4	3	7	8
		OC	Family	26	3	5	8	8
			Friend	33.3	6	5	7	8
			Phone	27.9	3	5	7	8
8	Yes	WC	Reading	0.5	2	2	2	2
			Clinician	1.3	1	2	2	2
			Phone	3.4	1	1	2	3
		OC	Family	1.9	4	4	2	2
			Friend	2.1	4	3	2	2
			Phone	1.9	2	3	2	2
9	Yes	WC	Reading	9.2	5	3	6	6
			Clinician	9.9	4	5	5	7
			Phone	11.5	3	4	5	6
		OC	Family	0.9	2	3	2	2
			Friend	3.8	4	4	3	6
			Phone	4.4	3	4	3	5
10	Yes	WC	Reading	0.7	2	1	2	3
			Clinician	0.9	2	2	2	2
			Phone	3.3	3	2	3	4
		OC	Family	0.3	1	1	2	2
			Friend	3.3	4	2	2	3
			Phone	2.6	3	5	2	5

Note. Ratings were made on a 9-point scale ranging from 1 *(least severe stuttering)* to 9 *(most severe stuttering)*. Previous therapy involved identification of individual stutters but not use of a rating scale. %SS = percentage syllables stuttered; immed = immediate; SLP = speech-language pathologist; WC = within the clinic; OC = outside the clinic.