

A Scale for Rating Clinician, Patient and  
Session Characteristics in Aphasia Treatment Sessions

R.H. Brookshire and L.E. Nicholas

Veterans Administration Medical Center, Minneapolis, Minnesota

During the past four years, we have developed a system for categorizing and recording events which occur within clinician-patient interactions in aphasia treatment sessions. Observers who use this system can reliably record almost all the events that occur within treatment sessions in considerable detail. The result of such recording is a quantitative summary of a given treatment session, in terms of the number of times various events and combinations of events occurred within the session.

However, as we gained experience with our coding system, it became apparent that, in many cases, such frequency-counts of events within sessions were not sufficient to describe those sessions. In many cases, sessions which contained similar proportions and distributions of events differed greatly in non-event-specific characteristics, such as clinician enthusiasm, clinician affect, patient affect, patient responsiveness, and so forth. Therefore, we designed a scale to measure these nonspecific clinician, patient, and session characteristics.

First, a list of potential descriptive categories was formulated by speech pathologists who knew our specific-event coding system, and who had extensive experience in observation of aphasia treatment. Then the proposed descriptive categories were evaluated by having two of these speech pathologists view a series of samples of aphasia treatment and rate each sample, using each descriptive category. When we identified potentially important treatment session, clinician, or patient characteristics that were not measured by a category on our list, we added a category to the list. Likewise, we eliminated what appeared to be redundant, overlapping, or trivial categories. When it appeared that the two judges could not reliably use a given category, the category was revised or eliminated. After a series of such trials, we generated the rating sheet shown in Table 1.

As can be seen in Table 1, the rating sheet contains 21 descriptors; 11 describe session characteristics, 5 describe clinician characteristics, and 5 describe patient characteristics. Each characteristic is rated on a five-point scale, with the opposite ends of the scale representing polar opposites. In general, "positive" ends of the scale are on the left, and "negative" ends are on the right.

In order to determine whether observers could reliably rate treatment sessions using this rating sheet, we asked 11 speech pathologists who were experienced in aphasia treatment to watch 40 ten-minute samples of aphasia treatment, and to rate each sample, using the 21 scales on the rating sheet. No formal definitions for any of the scale descriptors were provided; the 11 observers simply were instructed that the two ends of each scale represented the extreme values for each descriptor, and that they should rate each tape on each of the 21 scales, using whatever "definition" of each descriptor that they felt was appropriate. All 11 observers viewed and rated the tapes together, in two sessions, one week apart. Conversations about the tapes and comparisons of ratings during the observation sessions were not allowed.

Table 1. A scale used for rating clinician, patient and session characteristics in aphasia treatment.

SESSION

Goal Oriented	_____	Non-goal-Oriented
Organized	_____	Disorganized
Beneficial	_____	Nonbeneficial
Structured	_____	Nonstructured
Clinician Controlled	_____	Patient Controlled
Request Clear	_____	Ambiguous
Appropriate	_____	Inappropriate
Feedback Clear	_____	Ambiguous
Appropriate	_____	Inappropriate
Frequent	_____	Infrequent
Nonthreatening	_____	Threatening

CLINICIAN

Sensitive	_____	Insensitive
Supportive	_____	Nonsupportive
Formal	_____	Informal
Sincere	_____	Insincere
Enthusiastic	_____	Apathetic

PATIENT

At Ease	_____	Frustrated
Responsive	_____	Unresponsive
Happy	_____	Sad
Independent	_____	Dependent
Alert	_____	Confused

Most of our observers adopted the following strategy. First, they watched approximately five minutes of the sample, without making any marks on the rating sheet. At approximately five minutes, they marked (in pencil) their initial judgment for each of the scales. Then, during the last five minutes of the sample, they revised these initial judgments where they felt such revision appropriate. When the ten-minute sample ended, they were given approximately two minutes to complete final revisions of their judgments. Then the next treatment sample was played. (During the course of our sessions, we found it helpful to stop the videotape at approximately five minutes, and to give the observers about a minute to mark their initial judgments. By doing this, observers did not miss portions of the sample which occurred while they were marking their judgments on the rating sheet.)

In order to determine which categories could be reliably used by observers, we first determined that, given a five-point scale and eleven observers, if one looks at each scale, using elementary probability analyses, the probability that eight of eleven judgments will fall at two adjoining values by chance is less than .05. Accordingly, we used this criterion to evaluate whether each scale was reliable within a given treatment sample.

In order to determine the reliability of judgments for each scale across all 40 treatment samples, we used basic probability tables to determine that the probability that a scale would meet our earlier agreement criterion on 29 of the 40 samples by chance was less than 5%. Consequently, those scales for which agreement criteria were met on 29 or more of the 40 samples could be considered to be reliably judged, with the probability of error below .05. Figure 1 summarizes the results of these analyses.

The data contained in Figure 1 are encouraging, because they suggest that observers can reliably rate aphasia treatment sessions on a number of what might, at first glance, seem to be intuitive and subjective characteristics. We feel that reliability on those scales not presently meeting reliability criteria may improve if observers are given definitions of those scales. We have developed those definitions (Table 2).

In order to determine which of our rating scales might be related to other scales, we carried out a series of correlational analyses, in which we examined the correlations among all of our rating scale categories. These correlations were computed on all rating scale categories, whether or not they had been shown to be reliable. The discussion which follows will emphasize the reliable categories, although some interesting (to us) correlations (or non-correlations) involving nonreliable categories will be mentioned. In the tables which follow, nonreliable categories will be enclosed in parentheses.

Table 3 presents those significant correlations which involved session characteristics. Only two session characteristics were not involved in at least one significant correlation; "Feedback Frequent" and "Nonthreatening." Most of the session characteristics correlated highly with several other session characteristics. These high correlations among session characteristics may reflect causal mechanisms;—e.g., observers may rate a sample "goal oriented" because requests are clear and appropriate and feedback is clear and appropriate. On the other hand, these high correlations among session characteristics may reflect semantic overlap among categories; e.g., observers may have rated a sample "goal oriented" and "organized," and "beneficial" and "structured," because all these terms had similar meanings

RELIABILITY OF SUBJECTIVE RATING SCALE CATEGORIES

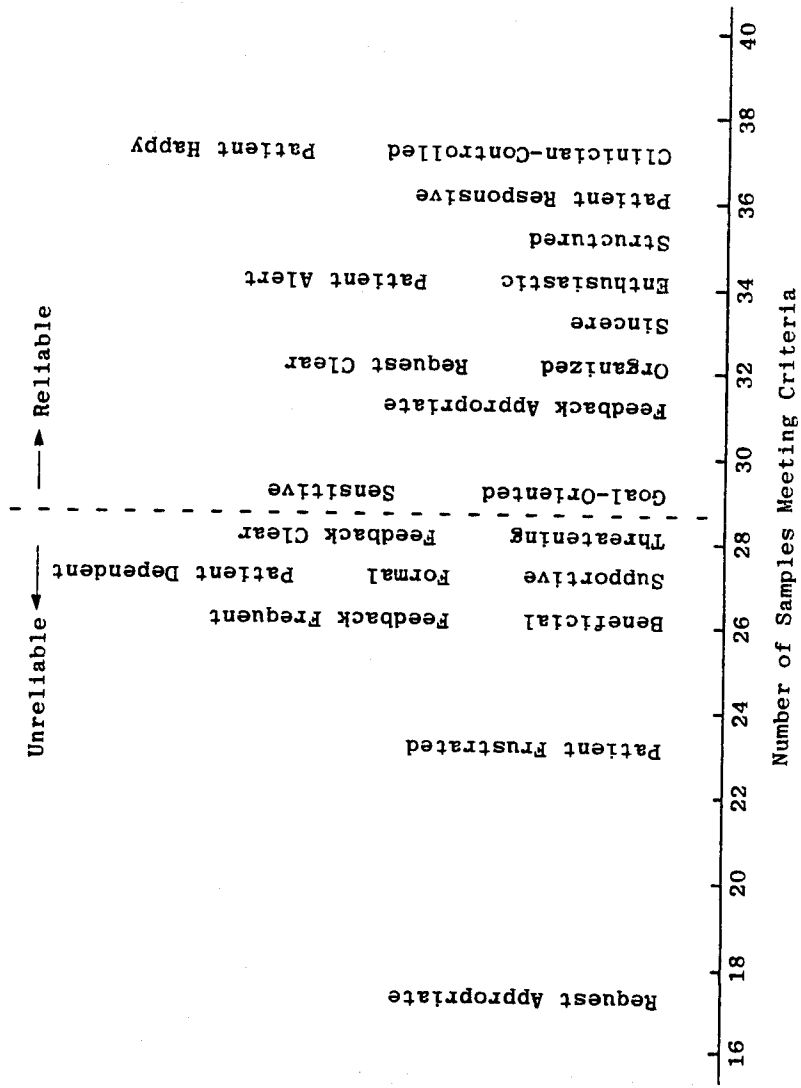


Figure 1. Reliability of subjective rating scale categories.

Table 2. Definitions for Subjective Rating Scale.

---

SESSION

Goal Oriented: Treatment activities are directed toward improvement of specific aspects of communication skills. Your rating of this aspect of the treatment session should reflect how easy it would be for you to write down one or more unambiguous goals, based on your observation of the session.

Beneficial: Treatment activities appear likely to result in improved communication abilities for the patient, either within the observed session or across a number of sessions.

Structured: Steps or stages within activities which occur in the session are arranged according to a pre-determined plan.

Organized: The session is conducted in a logical and orderly manner. Activities in the session appear to occur according to a systematic scheme, rather than occurring randomly.

Clinician Controlled: The clinician initiates and directs the interactions of the session. The patient's behavior has little effect upon the clinician's conduct of the session.

Request-Clear: The content of the clinician's requests is understandable, the intent is discernible, and the delivery is intelligible.

Request-Appropriate: The clinician's requests and other stimuli are appropriate to the patient's communication disabilities and performance level.

FB Clear: The content of the feedback delivered by the clinician is understandable and the intent is discernible.

FB Appropriate: The clinician's feedback is suitable for the patient's responses, for the treatment design, and for the patient's needs.

Nonthreatening: The clinician is creating a nonintimidating atmosphere for the patient.

CLINICIAN

Sincere: The clinician relates to the patient in a genuine and natural manner.

Supportive: The clinician assists and encourages the patient to perform, and is accepting of the patient's communicative attempts.

Sensitive: The clinician is aware of the patient's feelings, needs, and level of functioning and modifies his or her behavior, and the treatment session in response to the patient's feelings, needs, and level of performance.

Enthusiastic: The clinician displays enthusiasm for the patient and for the treatment activities.

PATIENT

Alert: The patient attends to the clinician and the treatment activities and is oriented to time, place, and circumstance.

Responsive: The patient responds consistently to the clinician's requests.

At Ease: The patient is relaxed and comfortable.

---

Table 3. Correlations among categories involving session characteristics.

<u>Goal-Oriented</u> correlated with:	
Organized	.87
(Beneficial)	.86
(Request Appropriate)	.85
(Feedback Clear)	.82
(Feedback Appropriate)	.79
Request Clear	.78
Structured	.76
*Sensitive	.73
*Sincere	.72
(*Supportive)	.69
<u>Organized</u> correlated with:	
Request Clear	.91
Structured	.89
Goal Oriented	.87
(Feedback Clear)	.86
Feedback Appropriate	.78
(Request Appropriate)	.71
<u>Beneficial</u> correlated with:	
(Request Appropriate)	.93
Goal Oriented	.86
*Sensitive	.82
*Sincere	.76
Feedback Clear	.75
<u>Structured</u> correlated with:	
Organized	.89
Request Clear	.77
Goal Oriented	.76
Feedback Appropriate	.74
(Feedback Clear)	.72
<u>Request Clear</u> correlated with:	
Organized	.91
(Feedback Clear)	.85
Goal Oriented	.78
Structured	.77
Feedback Appropriate	.70
<u>(Request Appropriate)</u> correlated with:	
(Beneficial)	.93
Goal Oriented	.85
*Sensitive	.83
Feedback Appropriate	.81
(Feedback Clear)	.79
*Sincere	.73
Organized	.71
(*Supportive)	.68
<u>Feedback Clear</u> correlated with:	
Feedback Appropriate	.88
Organized	.86
Request Clear	.85
Goal Oriented	.82
(Request Appropriate)	.79
(Beneficial)	.75
*Sensitive	.74
*Sincere	.73
Structured	.72
(*Nonthreatening)	.71

to our observers. At this time, we have not attempted to determine which of our categories are correlated in cause-effect fashion and which correlate because of such semantic similarity. In general, it seems likely that at least some of the correlations among categories under common headings (Session, Clinician, or Patient characteristics) may reflect semantic, rather than causal relationships. It also seems likely that correlations among categories which are not under common headings may be more likely to reflect causal than semantic relationships. Therefore, we shall give somewhat more attention to correlations between categories which are not under a common heading. These correlations are marked with asterisks in the tables.

Now, to return to Table 3, we can see that three "Clinician" categories were related to the "Goal Oriented" session category; "Sensitive," "Sincere," and "Supportive." As we shall see, these clinician categories appear frequently in other correlations involving session characteristics; namely, "Beneficial," "Request Appropriate," and "Feedback Clear" (Table 3). They did not correlate strongly with "Organized," "Structured," and "Request Clear" session characteristics. These data suggest that our observers saw "sincere," "sensitive," and "supportive" clinicians as providing "beneficial" treatment. Furthermore, it appears that "appropriate requests" and "appropriate feedback" were seen by our observers as characteristics of "sensitive," "sincere," and "supportive" clinicians. However, organization, structure, and clarity of requests were not seen by our observers as characteristics of "sensitive," "sincere," and "supportive" clinicians. It is also interesting (and probably important) that "Request Appropriate," "Sensitive," and "Sincere," together with "Feedback Clear" and "Goal Oriented" were strongly related to judgments of "Beneficial" by our observers.

Table 4 presents those significant correlations that involved clinician characteristics. As can be seen in Table 4, "Supportive," "Sincere," and "Sensitive" appear in most of these correlations, and the correlations are all with session characteristics. In fact, those correlations are substantially the same correlations that we described earlier (Table 3), except that they are organized according to clinician characteristics, rather than session characteristics. The only correlations that we have not previously mentioned are the correlations between "Nonthreatening" and "Feedback Clear," and "Nonthreatening" and "Feedback Appropriate." These relationships suggest that a primary determinant of whether a treatment session will be seen as threatening or not is the nature (and perhaps the amount) of feedback delivered by the clinician. In general, clinicians who deliver clear and appropriate feedback seem to be seen as less threatening than those who deliver unclear and/or inappropriate feedback.

Table 4 presents the only significant correlation observed involving "Patient" categories; a correlation between "Patient Responsive" and "Patient Happy." Although three of the five "Patient" scales yielded reliable judgments by our observers, none correlated significantly with either "Session" or "Clinician" characteristics. Another problem with the "patient" categories is that judges tended to rate most tapes at about the midpoint of the scale. Thus, even though judges agreed on many of the "patient" scales, they tended not to feel that patients differed very much on these characteristics. At this point, we feel that this "clumping" of judgments of patients' characteristics at the midpoint of the scales may be

Table 4. Correlations among categories involving clinician and patient characteristics.

<u>Sensitive</u> correlated with:		
(Supportive)	.90	
Sincere	.89	
(Nonthreatening)	.83	
(*Request Appropriate)	.83	
(*Beneficial)	.82	
(*Feedback Clear)	.74	
*Feedback Appropriate	.73	
*Goal Oriented	.72	
<u>(Supportive)</u> correlated with:		
Nonthreatening	.91	
Sensitive	.90	
Sincere	.88	
*Feedback Appropriate	.74	
*Goal Oriented	.69	
(*Request Appropriate)	.68	
		<u>Nonthreatening</u> correlated with:
		*Feedback Appropriate .77
		(*Feedback Clear) .71
		<u>Sincere</u> correlated with:
		*Feedback Appropriate .81
		(*Beneficial) .76
		(*Request Appropriate) .73
		(*Feedback Clear) .73
		*Goal Oriented .76
		<u>Patient Responsive</u> correlated with:
		Patient Happy .76



a result of the constraints placed upon the patient by most traditional treatment procedures; patients seem not to be permitted (or led) to demonstrate dependence, unresponsiveness, frustration, and so forth, in most treatment sessions.

To summarize the results of our correlation analyses, three major "clusters" of categories emerged. One cluster contained session characteristics—"Beneficial," "Goal Oriented," "Request Appropriate," "Feedback Clear," and "Feedback Appropriate." One might label this cluster "Session Clarity and Directionality." Another cluster contained clinician characteristics—"Sensitive," "Sincere," and "Supportive." We might label this cluster "Clinician Rapport." The third cluster contains the first two clusters, and represents the relationship between "clinician rapport" and "session clarity and directionality."

In addition to the rather undramatic showing of patient characteristics categories in our analyses, we were impressed by the failure of "Clinician Enthusiasm" to correlate with judgments of session characteristics (such as beneficiality). It appears that our observers agree that enthusiasm, in itself, is not as important to the treatment interaction as other clinician characteristics, such as sincerity, sensitivity, or supportiveness. Whether the aphasic persons in the treatment session feel the same way remains to be seen.

In summary, then, we have devised a system by which observers can rate "subjective" characteristics of aphasia treatment sessions with reasonable reliability. We have also demonstrated that observers see at least some of those characteristics as important to the effectiveness of the treatment sessions observed. Work needs to continue to improve the reliability of our rating scales, to examine further their interrelationships, and to determine whether observers' judgments about the importance of given categories are reflected in the aphasic persons's rate of recovery in treatment. It is our feeling that neither the Clinical Interaction Analysis System nor our Subjective Rating Scale are sufficient in themselves to describe aphasia treatment. Whether the two together are sufficient is a question that we hope to answer in the future.

Q: Since there is evidence that a five-point scale generates a "regression to the mean" more than a seven-point scale does, did you consider the possibility that a seven-point scale might have generated more dispersion in the ratings of patient characteristics?

A: When we began evaluating scales, we considered the possibility that seven might be better than five. But as we evaluated observer reactions to the scales, we decided that five was better, because observers were somewhat "put out" on being asked to rate a given characteristic on a seven-point scale, when we gave them no definition for the characteristic. They felt less "put out" in rating those characteristics on five-point scales. I suspect that if we were to offer definitions, and, perhaps, training on rating characteristics, then a seven-point scale might be better.

Q: Do you consider the intervals in your scales equal?

A: No.