Effects Of Clinician Request And Feedback Behavior On Responses Of Aphasic Individuals In Speech And Language Treatment Sessions

R. H. Brookshire
Linda S. Nicholas

Veterans Administration Hospital
Minneapolis, Minnesota 55417

There is a substantial body of evidence which suggests that the interaction between clinician and client in treatment sessions or interviews is a two-way process. That is, the behavior of the clinician (or interviewer) affects the behavior of the client (or interviewee), and vice-versa (Krasner, 1958; Amidon and Hough, 1967; Greespoon, 1968). The clinical process in speech pathology appears to be no exception. The expectation that the clinician's behavior will in some manner affect the client's behavior is fundamental to conceptions of the speech pathologist's role in treatment of communication disorders. Unfortunately, our knowledge of how and why the clinician's behavior may affect the client's behavior lags behind our conviction that the clinician's behavior does make a difference. Holland (1969), in summarizing the state of affairs in aphasia therapy comments:

> While language clinicians have been primarily responsible for producing elegant and useful tools to assess aphasic performance, we are lamentably short in producing a body of techniques to improve aphasic performance. We have produced lots of principles and many lists of do's and don'ts for ourselves, but these are general guidelines. The application of these guidelines is then left to the creativity of the individual clinician. (p. 3)

One "principle" in aphasia therapy that has received substantial experimental attention has been the principle that error rates should be kept low in aphasia treatment activities. A number of writers have suggested that treatment programs for aphasic individuals should be designed so that the occurrence of error responses by the aphasic individual does not exceed certain limits. Schuell, et al. (1965) emphasizes that the patient should receive controlled stimulation so that he or she is not forced to struggle for the correct response or to correct erroneous responses. Porch (1967) emphasizes that the clinician should work at levels at which the aphasic individual emits delayed or self-corrected responses, rather than outright errors. Brookshire (1973) supports Porch's contention, and adds that patients should rarely, if ever, be placed in tasks in which errors outnumber correct responses.

Brookshire (1972) demonstrated that the occurrence of errors by aphasic subjects in a confrontation naming task generated disruptive effects which often persisted to items following the error. That is, an error on a given item was frequently followed by an error on the subsequent item, even though that subsequent item had previously been shown to be easy-to-name for the patient. In a subsequent experiment, Brookshire (1976) demonstrated the same effects for aphasic subjects in a task in which they pointed to sequences of pictures named by the experimenter. Brookshire concluded from the two studies that errors committed by aphasic patients in speech and language tasks generate effects that are likely to interfere with performance on trials that follow the errors. He recommended that programs utilized in treatment of aphasic patients endeavor to keep error rates low.

The evidence suggests, then, that errors by aphasic individuals tend to precipitate errors on subsequent trials. However, we do not know whether these errors on subsequent trials are a function of some as-yet-undefined patient characteristics, or whether the clinician's behavior in some way promotes the occurrence of multiple errors. The studies cited also provide no information about the variables which may elicit the initial error in a sequence of errors, and also provide no information about those clinician behaviors which might function to prevent errors on subsequent trials, once an error has occurred. Because of these concerns, an investigation was conducted which attempted to determine (a) whether certain clinician behaviors and task characteristics are related to the occurrence of patient error responses in speech and language treatment sessions for aphasic individuals, and (b) Whether "errors generate errors;" that is, whether error responses tend to occur in clusters, rather than being distributed uniformly throughout the treatment session.

Methods.

Collection of Treatment Videotapes. Forty videotapes of aphasia treatment were collected from clinical facilities distributed throughout the United States. A non-systematically chosen ten-minute continuous sample was extracted from each of these tapes. These 40 ten-minute treatment samples were then coded in their entirety using the Clinical Interaction Analysis System (CIAS). The Clinical Interaction Analysis System (Brookshire, 1976) was developed at the Minneapolis Veterans Administration Hospital. It is a system for recording events which occur in clinician-patient interactions in aphasia treatment sessions. The CIAS contains 39 event categories among which events that occur within clinician-patient interactions can be apportioned.

Preparation of Master Coding Records. Four judges who were sophisticated in the use of CIAS, and who each had demonstrated inter-examiner category-by-category reliabilities above .90 (reliability = agreed-upon events ÷ total events) collaborated to produce a master coding record for each of the 40 ten-minute treatment samples. Two judges (depending upon availability) jointly observed a given treatment sample and, after discussion if necessary, decided (a) when each event in the sample occurred and (b) how each event

should be coded. If the two judges agreed upon the coding of an event, it was entered in the master coding record for the treatment sample. When the two judges could not reach agreement on how an event should be coded (this happened less than 5% of the time), a third judge (from the group of four) viewed the event in question, and made an independent coding decision. If that decision supported one of the other judge's decision, the event was entered in the master record based on the 2/3 agreement. On those infrequent occasions in which the third judge's decision did not confirm one of the judge's decisions, the three judges discussed the event until they agreed upon how it should be entered in the master record sheet. The event was then entered in the master record. In this manner, master records were generated for each of the 40 treatment samples. These records contained every event which occurred in each sample, coded with the 39-category CIAS.[1]

Quantification of Content of Treatment Samples. For each of the 40 master coding records, the number of occurrences of each of the 39 event categories was determined. The total occurrence of each event category within a treatment sample was divided by the total number of events which occurred within that treatment sample, to yield the proportion of total events within a sample which contained each event category. These proportions formed the data-base upon which subsequent analyses were carried out.

Identification and Transcription of Clusters of Unacceptable Responses. Within the master coding records for each of the 40 videotaped treatment sessions, each event which contained an unacceptable patient response (UR) was identified.[2] Then these UR events were transcribed onto another coding record sheet, along with those non-UR (acceptable response, or AR) events which preceded and followed the UR event. If more than one UR event occurred successively, or if a single non-UR event occurred between two UR events, then the "cluster" of UR (or UR-AR-UR) events was transcribed to the coding record sheet, along with the non-UR events which preceded and followed the cluster. Each event transcribed was described by the 39 event categories in the CIAS. The coding record sheets which resulted contained 672 UR request events, along with 306 AR events which preceded UR events and 306 AR events which followed UR events. Single AR events which occurred within

---

[1]The reliability of these master records has subsequently been confirmed by a series of trials in which 10 observers were trained to use the CIAS, and then coded portions of the 40 samples. Their mean category-by-category agreement with the master records consistently exceeded .90, and in no case did a single event-category agreement fall below .80.

[2]The mean observer reliability for "Unacceptable Response" across ten judges coding ten 10-minute treatment samples was .91 (agreements ÷ total events), suggesting that judgements of unacceptable responses contained in the master coding records were dependable.

strings of UR's were not counted as AR events preceding or following UR events, since they both preceded and followed UR events.

After the coding record sheets for these UR-AR sequences were completed, the proportional occurrence of each of the 39 event categories was computed for UR events, for AR-preceding-UR events, and for AR-following-UR events. In order to determine whether the proportional occurrence of each event category within UR, initial AR, and final AR events differed from expected proportions, the proportional occurrence of each event category within these UR and AR events was subtracted from its overall proportion occurrence in all 40 treatment samples. The same procedures were carried out for initial AR and final AR requests.

Results and Discussion.

The results of calculations described in the previous section are presented in Table 1. In order to determine which of the differences contained in Table 1 were meaningful, a Z-score procedure was carried out, in which the variance of event categories across all events in all 40 tapes was used to generate probability statements regarding the reliability of the differences observed. The resulting Z-scores are presented in Table 1. Those Z-scores which suggested reliable differences (p less than .01) are marked with an asterisk.

Acceptable Response Events Preceeding Unacceptable Response Events. Within AR events which preceded UR events, the following categories occurred significantly more often than would be expected, based upon the proportion occurrence of these categories across all 40 samples: (a) Explanation, (b) Presence of Feedback, (c) Spoken Feedback, (d) Gestural Feedback, (c) Positive Feedback, (f) Elaboration Feedback. Repeated Expected Response and Negative Feedback occurred significantly less often than would be expected.

Unacceptable Response Events. The following categories occurred significantly more often than would be expected within events which contained unacceptable responses: (a) Completion, (b) Question 2 (other than yes-no), (c) Repeated Expected Response, (d) Negative Feedback, (e) Correction Feedback. The following categories occurred significantly less often than would be expected: (a) Presence of Feedback, (b) Spoken Feedback, (c) Gestural Feedback, (d) Positive Feedback, (e) Repetition Feedback. UR events also contained significantly fewer clinician-emitted words than the average for all events within the 40 samples.

Acceptable Response Events Following Unacceptable Response Events. Within AR events which followed UR events, the following event categories occurred significantly more often than would be expected: (a) Explanation, (b) Repeated Expected Response, (c) Presence of Feedback, (d) Spoken Feedback, (e) Gestural Feedback, (f) Positive Feedback, (g) Repetition Feedback, (h) Elaboration Feedback. The following event categories occurred significantly less often than would be expected: (a) Clinician Discourse, (b) Long Expected Response, (c) Delayed Expected Response, (d) Negative Feedback.

Table 1. Percentage occurrence of event categories in 40 treatment samples, and in events containing unacceptable responses (URs), and non-UR events preceding and following UR events (* = p. < .01).

Category	40 Tape		AR Preceding UR		UR		AR following UR	
TYPE	Percentage	Std. error	Difference	z	Difference	z	Difference	z
Imperative	28.7	3.51	-6.69	-1.91	5.27	1.50	0.75	0.21
Model	7.3	1.83	-2.40	-1.31	2.20	1.20	-1.44	-0.79
Completion	8.7	2.04	-1.51	-0.74	8.69	4.26*	2.39	1.17
Yes-No Question	9.9	1.87	-0.35	-0.19	-2.51	-1.36	1.25	.67
Other Question	12.1	2.33	4.64	1.99	11.58	4.97*	-2.28	-0.98
Nonverbal	10.9	2.36	-1.38	- .58	-2.85	-1.21	-4.35	-1.84
Explanation	7.1	0.74	7.36	9.95*	----	----	7.96	10.76*
Clinician Discourse	7.8	0.93	0.74	0.80	----	----	-4.51	-4.84*
COMPLEXITY								
Inference	59.6	3.49	-1.39	-0.40	6.02	1.72	3.22	0.92
Number of words[1]	5.7	0.33	0.56	1.69	-1.55	-4.70*	0.41	1.24
MANNER								
Spoken	90.1	2.44	5.49	2.25	-3.33	-1.36	1.05	0.43
Gestural	33.1	3.17	5.43	1.71	-0.94	-0.30	6.63	2.09
Melodic	0.6	0.34	-0.23	- .68	----	----	----	----
MATERIALS								
Object-Picture	51.2	4.60	-1.69	-0.37	5.95	1.29	-10.37	-2.25
Written	11.2	3.06	-3.08	-1.01	-1.50	-0.49	-3.05	-0.99
EXPECTED RESPONSE								
Long	23.6	3.21	-2.93	-0.91	-4.53	-1.42	-8.98	-2.80*
Spoken	57.7	4.45	-2.76	-0.62	4.22	.95	-0.16	-0.04
Melodic	2.8	1.53	-1.36	-0.89	-1.87	-1.22	-2.76	-1.80
Gestural	41.6	3.70	-2.61	-0.71	-4.81	-1.30	0.46	0.12
Written	1.3	0.75	1.01	1.35	0.16	0.21	-0.45	-0.60
Repeated	24.9	2.09	-7.03	-3.37*	12.63	6.04*	33.53	16.04*
Delayed	3.1	0.94	0.18	0.19	-1.91	-2.03	-2.66	-2.83*
SUPPORT								
Unison Spoken	4.4	1.51	-0.87	-0.58	-1.20	-0.79	-0.86	-0.57
Unison Gestural	3.2	1.66	-2.15	-1.30	----	----	----	----
FEEDBACK								
Presence	48.1	2.75	9.13	3.32*	-26.42	-9.61*	10.70	3.89*
Spoken	45.2	2.72	9.29	3.42*	-25.23	-9.28*	9.25	3.40*
Gestural	25.3	1.87	6.12	3.28*	-13.58	-7.26*	11.39	6.04*
Positive	42.9	2.85	12.05	4.23*	-30.52	-10.71*	8.89	3.12*
Negative	3.1	0.56	-2.21	-3.95*	6.38	11.39*	-2.70	-4.82*
Correction	1.7	0.29	-0.34	-1.17	3.61	12.45*	-0.42	-1.45
Repetition	7.4	1.13	2.88	2.55	-5.21	-4.61*	5.83	5.16*
Elaboration	4.7	0.69	2.83	4.10*	-0.96	-1.39	4.16	6.03*
Intense	0.6	0.28	0.34	1.21	-0.45	-1.61	-0.59	-2.11
Number of words		0.08						

1. Number of words = mean number of words per event

Interpretation of Event Category Differences. It seems reasonable to expect that the nature of a clinician's request might affect the patient's response to that request. One might also expect that, within a treatment session, certain kinds of clinician behaviors or certain types of requests might have the capacity to "set the stage" for patient errors to subsequent requests, even though they do not immediately elicit an error response. If certain clinician behaviors or requests introduced uncertainty or were confusing to the patient, or in some way introduced "noise" into the patient's language processing system, then we might speculate that the effects of these clinician behaviors could be delayed, so that they generated errors, not in the event within which they occurred, but in the next subsequent event. We found no evidence to support this speculation. The only request type which occurred with unusual frequency in AR events which preceded UR events was "Explanation". This suggests that patients make errors following explanation events. Since explanation is frequently used by clinicians at the outset of treatment activities, or at the transition from one activity to another, it appears that what we are seeing is a tendency for patients to make errors on initial items in treatment activity -- a tendency which has been documented in previous literature (Brookshire, 1973).

We had expected that, if any event categories within AR events which preceded UR events were to have an effect upon responses to subsequent clinician requests, then feedback was the most likely candidate. It seemed reasonable to expect that if the clinician's feedback for a patient response was inappropriate or confusing, or if the clinician failed to provide feedback, then the patient's response to the following request might be adversely affected. This expectation was not supported by the results of our analyses. The occurrence of positive feedback and the general proportion of feedback on AR events which preceded UR events was significantly above the overall proportions of these event categories across the 40-tape sample, and the presence of negative feedback was significantly below the proportion for the 40-tape sample. These results suggest that unacceptable responses do not usually occur as the result of clinicians' failure to provide appropriate reinforcement for preceding responses.

The significantly higher-than-expected-proportion of Repetition and Elaboration feedback in these pre-UR events might suggest that if the clinician repeats or elaborates upon the patient's response, then the next patient response is more likely to be unacceptable then we would otherwise expect. However, Repetition and Elaboration feedback also occurred more frequently than predicted in AR events which followed UR events, suggesting that clinicians may have a general tendency to repeat and elaborate upon any acceptable response. The lower-than-expected proportions of Repeated Expected Responses in AR events which preceded UR events appears to be a result of the tendency for clinicians to limit the use of RER to events which follow unacceptable responses.

There was a strong tendency for Completion and Question 2 (other than yes-no) to occur within requests which contained unacceptable responses. This suggests either that Completion and Question 2 requests may be more difficult for aphasic patients, or that one or both may be used in circum-

stances in which error rates are high, such as when the patient is severely aphasic, and can perform few of the "traditional" treatment tasks with consistent success. Our impression is that Question 2 requests are, indeed, difficult for aphasic patients, and that they tend to elicit errors. On the other hand, our observations of the treatment samples suggest that Completion requests tend to occur in situations where the patient is having difficulty, and that Completion requests tend to be used in treatment for patients with severe deficits -- especially those with severe formulation and speaking problems.

The results of this study suggest that clinicians tended to respond to unacceptable responses in a characteristic manner. There was a strong tendency for clinicians not to provide feedback for unacceptable responses. The presence of Feedback, Spoken or Gestural, within events containing unacceptable responses was far below expected feedback proportions. However, negative feedback and correction feedback occurred with higher-than-expected frequency within UR events, suggesting that when feedback for unacceptable responses does occur, it is likely to be negative and contain correction. (But not repetition, which occurred less often than would be expected with UR events.) The presence of higher-than-expected proportions of Repeated Expected Responses within UR events may be related to a tendency for clinicians to respond to patient errors by asking for the same response again, along with the tendency for UR's to cluster. That is, if the patient makes an error, then the clinician is likely to ask for the same response again in the next event. Because errors tend to cluster, the patient's response to that request is likely to be unacceptable. This seems to explain the tendency for UR events to contain Repeated Expected Responses in greater-than-expected proportions.

A somewhat surprising result was the finding that requests which generated unacceptable responses were significantly shorter than the average length for all events across the 40 tapes. One might logically expect that longer requests would be more likely to generate unacceptable responses than shorter requests, but, according to our analysis, UR events contained significantly fewer words than one would expect, based on the 40-tape mean length of clinician utterance. However, these utterances contained Explanation and Clinician Discourse, as well as request utterances, and we might expect that Explanation and Clinician Discourse might tend to contain more words than request events do. Therefore, we tabulated the mean number of words contained within a sample of non-UR request events which neither preceded nor followed UR events, to obtain as estimate of event length for request events that did not generate or follow unacceptable responses. The mean length of these request events was 4.73 words. A test on the difference between the length of these non-UR (4.73 words) and our previously-tabulated UR request events (4.17 words), yielded a value of 1.35 (p. greater than .05) suggesting that there was no significant difference between the number of words in UR and non-UR request events.

The reluctance of clinicians to provide feedback for unacceptable responses does not extend to acceptable responses which follow unacceptable responses. As in the case of AR events which precede UR events, Positive,

Spoken and Gestural Feedback occurred with greater-than-expected frequency in AR events which followed unacceptable responses. There was also a strong (and highly significant) tendency for clinicians to provide Repetition and Elaboration feedback to these responses. Clinicians did not, of course, provide negative feedback to these acceptable responses. There was a greater-than-expected proportion of Explanation events within AR events which followed UR events, as there has been within AR events which preceded UR events. Observation of the 40 tapes suggested that pre-and post-UR events might contain two different kinds of explanation. It appeared that explanation which appeared within pre-UR events was generally instructional and descriptive of upcoming treatment activities, while explanation which followed UR events was generally corrective. For example, a typical pre-UR explanation might be, "Now I'm going to say the name of each of these pictures, and I want you to point to the one I name." A typical post-UR explanation might be, "Point to the pictures. Don't say their names." An analysis of pre-and post-UR explanation events confirmed this hypothesis. Of 30 Explanation events which occurred in pre-UR events, 25 were instructional, 13 were corrective, and one was unclassifiable. Of 60 Explanation events which occurred in post-UR events, 47 were corrective and 13 were instructional. Both differences were significant (Z = score approximation to the binomial distribution, p. less than .05), supporting the hypothesis that pre-UR Explanation events were primarily instructional, and post-UR events were primarily corrective.

Clustering of Unacceptable Responses. In order to determine whether events containing unacceptable responses tended to cluster, the precentage of UR events was computed for each of the 40 tapes. For each tape, the percent of UR events which occurred in groups of two or more consecutive UR events was then computed. Then, for each tape, the percent of UR's in the entire tape was subtracted from the percent UR's in clusters. If there were no tendency for UR's to cluster, the two percentages should be equal or nearly equal. If there were a tendency for UR's to cluster, a positive difference should be observed. There were 25 positive differences (Mean = 20.57, SD = 11.73, Range = 2.96-52.09), and 13 negative differences (Mean = 5.40, SD = 3.84, Range = 1.30-16.80). Fourteen of the 25 positive differences were greater than any of the negative differences. (There were two ties.) The differences were then arranged in rank-order, and a Wilcoxen matched-pairs-signed-ranks test were performed. The obtained Z was 3.71 (p. less than .01), which indicated that more UR's occurred adjacent to another UR events than could be accounted for by chance, and confirmed that unacceptable responses tended to occur in clusters.

Summary. Records of the content of forty videotaped samples of aphasia treatment sessions were prepared. These records were analyzed to determine 1) whether certain clinician behaviors and task characteristics are related to the occurrence of patient error responses in speech and language treatment sessions for aphasic individuals and 2) whether "errors generate errors"; that is, whether error responses tend to occur in clusters rather than being distributed uniformly throughout the treatment session. The results of these analyses indicated that a number of event categories were significantly related to the occurrence of unacceptable patient responses. These results suggest that certain clinician behaviors generate patient error responses and also that clinicians tend to respond to patient errors in characteristic ways. An analysis of these results also confirmed that unacceptable patient responses tended to occur in clusters.

References:

Amidon, E.J., Hough, J.B. (eds.), Interaction Analysis: Theory Research, and Application. Reading, MA: Addison-Wesley Publishing Co., 1967.

Boone, D.R., Prescott, T.E., Audiotape and Videotape Self-Confrontation Manual. Denver, CO: University of Denver, 1970.

Brookshire, R.H., Effects of task difficulty on naming performance of aphasic subjects. J. Speech Hear. Res., 15, 551-558, 1972.

Brookshire, R.H., The use of consequences in speech pathology: Incentive and feedback functions. J. Commun. Dis, 6, 88-92, 1973.

Brookshire, R.H., An Introduction to Aphasia. Minneapolis, MN: BRK Publishers, 1973.

Brookshire, R.H., Effects of task difficulty on sentence comprehension perfo4mance of aphasic subjects. J. Commun. Dis., 9, 167-173, 1976.

Brookshire, R.H., A system for coding and recording events in patient-clinician interaction during aphasia treatment sessions. In: R. Brookshire (ed.), Clinical Aphasiology, Minneapolis, MN: BRK Publishers, 1976.

Diedrich, W.M., Charting Speech Behavior. Lawrence, KS: University of Kansas, 1973.

Greenspoon, J., Simkins, L., A measurement approach to psychotherapy. Psychological Record, 18, 409-423, 1968.

Hively, W., Programming stimuli in matching to sample. J. Exp. Anal. Behav. 5, 279-288, 1962.

Holland, A.L., Some current trends in aphasia rehabilitation. ASHA, 11, 3-7, 1969.

Johnson, T.S., The development of a multidimensional scoring system for scoring the clinical process in speech pathology. Unpublished Ph.D. dissertation, Lawrence, KS: University of Kansas, 1969.

Krasner, L., Studies of the conditioning of verbal behavior. Psych Bull. 55, 148-169, 1958.

Oration, A.R., Supervision in Speech Pathology: A Handbook for Supervisors and Clinicians. Baltimore, MD, 1977.

Porch, B.E., Porch Index of Communicative Ability. Palo Alto, CA: Consulting Psychologists Press, 1967.

Schubert, G.W., Miner, A., Modification of the Flanders interaction analysis categories for observation in speech therapy. Paper presented at American Speech and Hearing Association Convention, Chicago, IL, 1971.

Schuell, H., Jenkins, J., Jimenes-Pabon, E., Aphasia in Adults: Diagnosis, Prognosis, and Treatment. New York, NY: Harper and Row, 1965.

Stoddard, L.T., Sidman, M., The effects of errors on childrens performance on a circle-ellipse discrimination. J. Exp. Anal. Behav. 10, 261-270, 1967.

Taylor, M.L., Language Therapy. In: H.G. Burr (ed.), The Aphasic Adult: Evaluation and Rehabilitation. Charlottesville, VA: Wayside Press, 1964.

Terrace, H.S., Discrimination learning with and without "errors." J. Exp. Anal. Behav., 6, 1-27, 1963.

Touchette, P.E., The effects of graduated stimulus change on the acquisition of simple discrimination in severely retarded boys. J. Exp. Anal. Behav., 11, 39-48, 1968.