CHAPTER

25

Conducting Research in a Clinical Setting Against All Odds: Unusual Treatment of Fluent Aphasia

James G. Schumacher
Linda E. Nicholas

Studies to evaluate treatment approaches for improving the speech production of patients with fluent aphasia are somewhat lacking in the literature. When we were confronted with a patient with a fluent aphasia and concomitant oral, nonverbal apraxia, we attempted a treatment approach commonly used for fluent aphasia. When this did not appear to be effective, we decided to examine whether a treatment commonly used for apraxia of speech would be effective in improving the patient's severely paraphasic output. We were additionally interested in examining the problems one encounters when attempting to conduct research in a clinical setting.

METHOD

SUBJECT

Mr. H was a 72-year-old man who sustained a single left-hemisphere thromboembolic CVA in October of 1987. Approximately 10 months following his CVA, Mr. H requested speech and language treatment at our medical center. His conversational speech was fluent but contained frequent literal paraphasias, as well as some verbal paraphasias and islands of effortless, correctly produced speech. Neologisms were frequent, particularly in more structured interactions and tasks. Mr. H also displayed a moderate oral, nonverbal apraxia. Prior to initiation of this study, Mr. H's aphasia quotient on the Western Aphasia Battery (WAB) (Kertesz, 1982) was 57.5, with a profile characterized by a fluent output that relayed a moderate amount of information, a mild to moderate auditory comprehension deficit, and severe deficits in repetition and naming.

STIMULUS MATERIALS AND PROCEDURES

Sixty black-and-white line drawings of high-frequency nouns of an original set of 95 met our criterion of no correct productions across three baseline sessions. Forty nouns were monosyllabic and 20 were bisyllabic. Treatment was provided on 20 randomly selected monosyllabic words from the set of 40. A combined ABA single-subject multiple-baseline design across behaviors was used to assess treatment effects. Probe measures were obtained at the beginning of every other treatment session on 10 treated and 10 untreated monosyllabic words. There were two sets of 10 treated and 10 untreated monosyllabic words (lists A and B). These were probed alternately. Probe measures also were obtained at the beginning of every ninth treatment session for alternating lists of 10 untreated bisyllabic words (lists A and B). Baseline and test measures (Table 25-1) included confrontation naming of all 60 line drawings, part F of the Token Test from the Neurosensory Center Comprehensive Examination for Aphasia (NCCEA) (Spreen and Benton, 1969), a writing to dictation task for five words, and a connected-speech task (two picture descriptions and two procedure descriptions). All test measures were administered after every 16 treatment sessions. All baseline, test, and probe measures were administered by the first author and scored by the second author. Confrontation naming of the line drawings was scored plus or minus, and the score was determined from the subject's first complete response. Writing words to dictation was scored plus or minus also. The number of correct information units (CIUs) (Nicholas and Brookshire, 1988) was scored for each connected-speech sample (picture and procedure descriptions). CIUs are informative words that are intelligible in context and accurately communicate information relevant to the picture or procedure.

Interjudge and intrajudge reliability was computed for all measures used in baseline, tests, and probes. Reliability was computed for all measures in 1 of the 3 baseline sessions, 2 of the 6 test sessions, and 6 of the 19 probe sessions. All point-to-point reliability values were at or above 87 percent and ranged from 87 to 100 percent.

Treatment was provided four times weekly, with two 45-minute sessions on each of 2 days (one session before lunch and one after lunch on each treatment day). All treatment was provided by the first author. Treatment tasks were chosen based on Mr. H's response to various techniques used to improve production of single words. As displayed in

TABLE 25-1. BASELINE AND TEST MEASURES

1. *Word retrieval:* Confrontation naming of pictures (40 monosyllabic and 20 bisyllabic)
2. *Auditory comprehension:* NCCEA Token Test, Part F
3. *Writing:* Words (man, has, rain, watch, yellow) written to dictation
4. *Spontaneous speech:*
   a. BDAE cookie theft picture
   b. WAB picnic picture
   c. Procedure description:
      (1) Describe how you would change a tire
      (2) Describe how you would wash a car by hand

Note: Half the monosyllabic words were treatment stimuli; NCCEA = Neurosensory Center Comprehensive Examination for Aphasia (Spreen and Benton, 1969); BDAE = Boston Diagnostic Aphasia Examination (Goodglass and Kaplan, 1983); WAB = Western Aphasia Battery (Kertesz, 1982).

Appendix 25A, treatment tasks were arranged in a hierarchy of steps of increasing difficulty for this patient. The hierarchy consisted of combined auditory (imitation) and visual (reading) stimulation, with subsequent fading of auditory and then visual stimuli at successive steps of the hierarchy. These tasks were similar to those commonly used in treatment for patients with paraphasic output. Within the hierarchy, Mr. H progressed from step to step, with backups to a previous step or steps if failure occurred, so as to provide success at the final step if possible. Training criterion was set at 80 percent accuracy across three consecutive sets. Responses made during treatment were scored as plus or minus. An error response was one in which a literal or verbal paraphasia, neologism, self-correction, or rejection response was produced. Delayed responses and distorted responses that did not cross phoneme boundaries were scored as correct.

Problem

Even though our probe data suggested that Mr. H's production of monosyllabic words was improving, his performance during treatment sessions was actually deteriorating. We felt that this might be due to two factors. First, the combination of input modes chosen (auditory, written) produced poorer performance than a single-channel stimulus. Second, working up the hierarchy until Mr. H made an error and then providing backups until he was once again successful appeared to make his performance worse. For example, if he was successful at steps 1, 2, and 3 but failed on step 4, he then also might fail at steps 3 and 2 on backups, leaving his final successful level at step 1 rather than his initially successful level of 3.

Solution

We constructed a new hierarchy based on clinical observations made during the first treatment. We observed that if Mr. H could successfully produce the first sound in a word, he would often produce the entire word correctly. In addition, his moderate oral, nonverbal apraxia became more evident when single-word stimuli were treated intensively, and this appeared to interfere with his ability to produce the first sound of a word. Consequently, we decided to treat his speech-production problems with a hierarchy of tasks traditionally used for apraxia of speech. This second treatment hierarchy, displayed in Appendix 25A, also consisted of four steps, with step 1 now consisting of combined auditory cues (imitation and verbal-placement information) and tactile cues (direct placement or traditional motokinesthetic). Steps 2 and 3 of the hierarchy then faded tactile cues and imitation, respectively. Verbal-placement cues were eliminated in step 4 of the hierarchy. Verbal-placement and tactile cues were given only for the initial sound of each stimulus item. Training criterion was again set at 80 percent across three consecutive treatment sessions.

Problem

The problems encountered with the initial cueing hierarchy were again encountered. That is, Mr. H's final successful level was often lower than his initial successful level before backups.

Solution

The second treatment hierarchy was continued. However, training was modified to be given only at each step until training criterion was reached at that step. Once step 1 was trained to criterion, step 2 was begun. As such, the treatment hierarchy itself resembled a component-assessment design with sequential deletions of components within the treatment package.

Problem

Sensitivity of Scoring. Although Mr. H continued to show positive changes on probe data, his plus and minus scores did not reflect improvement to the extent that we were noting it clinically. For example, in the beginning of treatment, he frequently produced a neologism for a given word, but as treatment progressed, he more closely approximated target words and frequently was only one phoneme away from the target. Both neologisms and closed approximations would receive a minus score using the plus or minus scoring system.

Solution

The second author rescored all probes using a modified 15-point PICA scoring system. Both plus or minus and PICA scores for probe data are reported in Table 25-2. When plus or minus scoring of probes was used, the difference between the mean for the three pretreatment baseline probes and the final four probes for treated words was substantial. Also, modest generalization of treatment effects was noted to nontreated monosyllabic words. Treatment effects were maintained at a fairly high level both immediately following withdrawal of treatment and at 3 months after treatment. When monosyllabic word responses were rescored using PICA scoring, the positive effects of treatment were more evident than with plus or minus scoring for both treated and untreated words.

TABLE 25-2. MEANS FOR TREATED AND UNTREATED MONOSYLLABIC WORD LISTS USING THREE MEASURES OF PERFORMANCE

	Baseline (three sessions)		*Treatment (last four sessions)*		*Follow-up (four sessions)*	
	TREATED	UNTREATED	TREATED	UNTREATED	TREATED	UNTREATED
Percent correct (+/−)	0	0	52.5	25.0	40.0	30.0
PICA score	4.5	4.4	9.2	7.2	8.6	7.5
Percent jargon	81.7	86.7	40.0	47.5	40.0	52.5

Finally, since jargon responses were abundant at the beginning of treatment, we were interested in the percentages of jargon responses produced throughout treatment and at follow-up testing. Our operational definition of a *jargon response* was one in which Mr. H's production was more than one phoneme away from the target word. The percent of jargon responses decreased noticeably from the pretreatment baseline probes to the final four probes during treatment for both treated and untreated words (Table 25-2). The decrease in jargon responses was maintained following withdrawal of treatment. This analysis suggests that even though many responses still were not correctly produced, a larger percentage of them were only one phoneme away from the target word than at the beginning of treatment.

Table 25-2 presents data for monosyllabic words only. As seen in Figure 25-1, performance on untreated bisyllabic words, picture and procedure descriptions, Token Test Part F, and writing to dictation remained essentially unchanged or fluctuated throughout treatment, with no distinct trend for improving or declining performance.

Of the four steps within the second or placement hierarchy, Mr. H had reached criterion on steps 1 and 2 and was nearing criterion on step 3 at the conclusion of treatment. A total of 32 sessions of placement hierarchy treatment were provided. Treatment was discontinued at that time because changes, although positive, were occurring very slowly and were not generalizing in a consistent way to the connected-speech samples in the study.

DISCUSSION

During and upon completion of this investigation, we were confronted with six primary questions we feel need to be addressed when attempting to conduct similar research in a clinical setting.

1. Should decisions about the effectiveness of a treatment be based on treatment data or probe data? We chose to discontinue the first treatment hierarchy because of Mr. H's deteriorating performance in treatment and his frustration exhibited during the treatment, even though our probe data showed improving performance.

2. Can an ineffective treatment be modified during a study without invalidating the study? The answer to this is "yes" if the modifications are carried out systematically. Treatment designs such as component assessment (additive or reduction) can be implemented either a priori or during the study to examine which components or combinations of

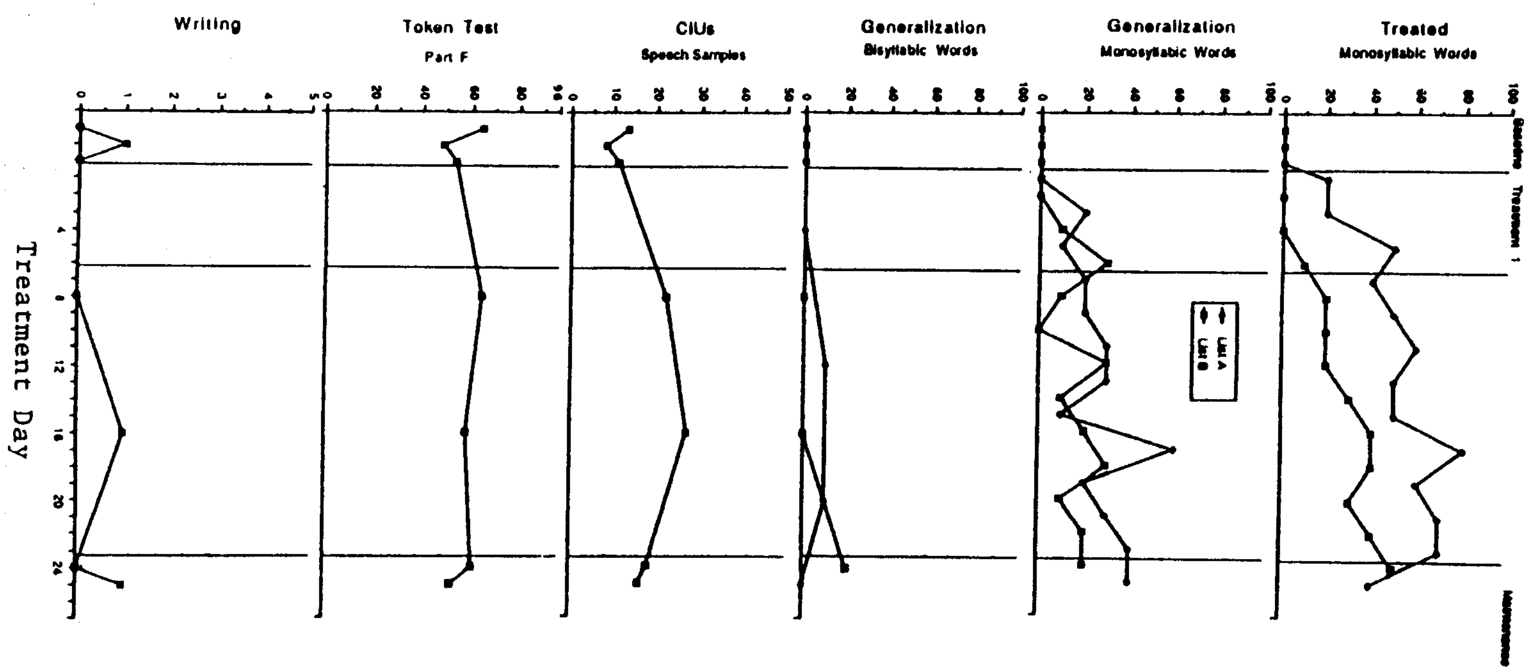

Fig. 25-1. Baseline, probe, and maintenance data for treated and generalization stimuli. Lists A and B = split lists of monosyllabic and bisyllabic words used alternately in probes. CIUs in speech samples = mean number of correct information units for picture and procedure descriptions combined. Token Test, Part F = 96 total points possible. Writing = number of words correctly written to dictation of five possible.

components of a treatment package are effective. Our modifications were carried out in only a partially systematic manner, but other reports suggest that systematic modifications that do not jeopardize the study are possible. Such modifications are discussed in excellent detail in a trio of articles on conducting single-subject research by McReynolds and Thompson (1986), Kearns (1986), and Connell and Thompson (1986).

3. Should the nature of the desired change dictate the sensitivity of the scoring system used? It is likely that the nature of the aphasia may dictate whether changes in performance will occur in an all-or-nothing fashion or in a stepwise fashion (e.g., from neologism to intelligible paraphasia to correct production). An appropriate scoring system could be chosen based on these predictions.

4. If hierarchies are used, should performance be trained to criterion at each step before proceeding to the next step? We feel that there are patients who tolerate extended drill on individual stimulus items and those who do best if the clinician moves on once a desired response is obtained. The decision to train hierarchies in their entirety or train to criterion step by step probably should be made based on subject variables. Likewise, perhaps generalization should be programmed at each step rather than at the final step of a hierarchy. It seems possible that failure to generalize may be due to programming of generalization too late and at a level too difficult for the patient to tolerate.

5. Can treatment studies be designed to allow for flexibility and maximum interest level for the clinician and patient? Stokes and Baer's (1977) description of "loose training" seems particularly appropriate for addressing this question. By using multiple clinicians, different environments, and a large number of treatment stimuli, the clinician and patient are not subjected to endless and often tedious drills on a small set of stimuli, and there may be a greater probability that generalization of treatment effects will occur.

6. And finally, is the collection and analysis of baseline, probe, and generalization data (with reliability checks) feasible in a clinical setting? Certainly extensive collection and analysis of baseline, probe, and generalization data is time-consuming. Realistically, it is our feeling that research of this type cannot occur unless the clinician's job allows time for conducting research, or unless support staff are available for collecting and analyzing data.

In conclusion, we learned a great deal from this effort. We now realize what a time commitment such studies require. We also understand the importance of allowing for flexibility in the treatment design both to allow for modification of an ineffective treatment and to program for generalization. Although we are unable to ascertain which components were responsible for the treatment effects noted in our treatment package, we were able to demonstrate improved performance, with generalization and maintenance of treatment effects. Hindsight is admittedly 20-20, and we fully acknowledge the problem areas that we discussed but did not address in this study. We feel more confident that research can occur in select clinical settings without compromising the clinician's intuitions about optimal treatment, the flexibility of the treatment itself, and the potentially healthy relationship between researcher and clinician.

REFERENCES

Connell, P. J., and Thompson, C. K. (1986). Flexibility of single-subject designs: III. Using flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders, 51,* 214–225.

Goodglass, H., and Kaplan, E. (1983). *The Boston diagnostic aphasia examination.* Philadelphia: Lea & Febiger.

Kearns, K. (1986). Flexibility of single-subject designs: II. Design selection and arrangement of experimental phases. *Journal of Speech and Hearing Disorders, 51,* 204–213.

Kertesz, A. (1982). *Western aphasia battery.* New York: Grune & Stratton.

McReynolds, L. V., and Thompson, C. K. (1986). Flexibility of single-subject experimental designs: I. Review of the basics of single-subject designs. *Journal of Speech and Hearing Disorders, 51,* 194–203.

Nicholas, L. E., and Brookshire, R. H. (1988). Manual for scoring correct information units (CIUs). Unpublished manuscript, Minneapolis, Minnesota.

Spreen, O., and Benton, A. L. (1969). *Neurosensory center comprehensive examination for aphasia.* Victoria, British Columbia: Department of Psychology, University of Victoria.

Stokes, T. F., and Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367.

APPENDIX 25A. TREATMENT HIERARCHIES

First treatment hierarchy:

Step 1. Patient produces the target word after he hears a model from the clinician (imitation) and sees the written word (reading).
Step 2. Patient produces the target word after he sees the written word.
Step 3. Patient produces the target word after he sees the initial letter of the target word.
Step 4. Patient produces the target word with only a picture prompt.

Second treatment hierarchy:

Step 1. Patient produces the target word after he hears a model (imitation), is given verbal placement instructions, is provided with direct tactile placement cues, and hears the model again.
Step 2. Patient produces the target word after he hears a model and is given verbal placement instructions.
Step 3. Patient produces the target word after he is given verbal placement instructions.
Step 4. Patient produces the target word with only a picture prompt.

Note: Pictures were visible to the patient at all steps in both hierarchies.