Efficacy During Acute Rehabilitation

Reg L. Warren, Cynthia Gabriel, and Ann Johnston
Braintree Hospital, Braintree, Massachusetts

Ann Gaddie
Braintree Hospital and Veterans Administration Hospital, Boston, Massachusetts

Single subject experimental design has led to numerous advances in our attempts to study the efficacy of treatment of aphasic patients. For good reasons, most studies have investigated the treatment of patients beyond the period of spontaneous recovery. This paper describes our initial attempts at Braintree Hospital to study the efficacy of treatment in a for-profit, acute rehabilitation setting. Typically, our patients are rehabilitated between two and ten weeks post onset and have an average length of stay of 38 days. They receive 30-35 half-hour treatment sessions. To date we have conducted twelve studies utilizing multiple baseline designs across behaviors to study treatment--no treatment comparisons. The average length of our studies is 12 treatment sessions across 15 calendar days. The following is a summary of three such studies:

Patient #1. C.S. was a 65-year-old male with a left thalamic infarct and attenuation in the posterior temporal and occipital regions with a moderate anomic aphasia and an overall PICA percentile of 54. He was treated for 37 sessions between 2.5 and 6.5 weeks post onset. The study was conducted over 17 days between 4 and 6.5 weeks post onset during 14 sessions.

Table 1. Treatment program for Patient C.S.

CLINICAL GOAL:	Improve verbal labelling
STIMULI:	Two sets of ten ADL objects and "Gesture" card
PRE-Rx TRAINING:	Gesture object function with "Gesture" card cue
TREATMENT:	1) Name object 2) Orient to "Gesture" card 3) Clinician requests gesture 4) Clinician requests verbal description 5) Clinician describes attributes and function 6) Initial phoneme cue
PROBE:	Name object
SCORING:	Plus-minus (self-correction allowed)
RELIABILITY:	Across 40 probes: 100%
CRITERION:	80%
PRE-POST:	Confrontation-responsive naming (BDAE) Concept/syllable ratio

The clinical goal was to increase verbal labelling. The stimuli were two sets of 10 objects common to activities of daily life (ADL) and the printed word "gesture." In pretreatment training C.S. was taught to gesture object function when cued with a "gesture" card. A 6-step treatment hierarchy began with a probe level request to name the stimulus objects. If unable to do so, C.S. was oriented to the gesture card in order to elicit the gesture and hopefully the object name. If not, the clinician requested C.S. to gesture object function and then say the name. If not, the clinician requested a verbal description of the object. If this failed, the clinician described the object's attributes and function and requested the name. Finally, an initial phoneme cue was provided which always elicited the target response.

The probe was conducted prior to each treatment session. C.S. was asked to name each object. Order of lists probed was counterbalanced across sessions. Plus-minus scoring was used (self-corrections allowed). Reliability across 40 probes was 100%. Treatment was to be discontinued at 80%. Two BDAE subtests assessing naming and the concept/syllable ratio were selected for pre-post measures.

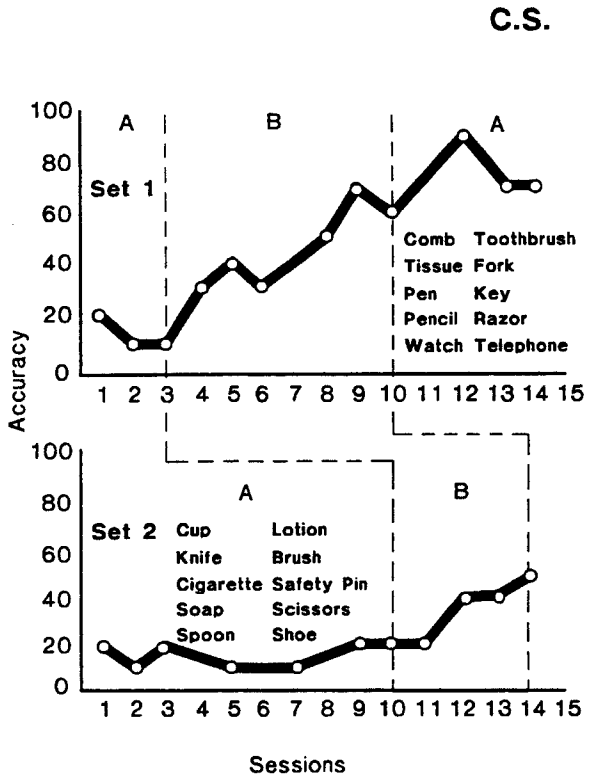

Figure 1. Performance of Subject C.S. on program to improve verbal labeling.

A multiple baseline design across two sets of objects showed a descending trend in baseline for Set 1 with improvement over 7 treatment sessions to a maximum of 70% accuracy (Figure 1). With discharge approaching in 4 or 5 days, the decision was made to initiate a maintenance phase for Set 1, which shows stabilization at 70%. Concurrently 7 probes of Set 2 performance remained low and relatively stable in baseline across 10 sessions. With the initiation of treatment, performance improved from 20 to 50% in 4 sessions followed by the patient's discharge. The difference in trend between the treatment phase of

Set 1 and the baseline phase of Set 2 is offered as evidence of a reliable treatment effect.

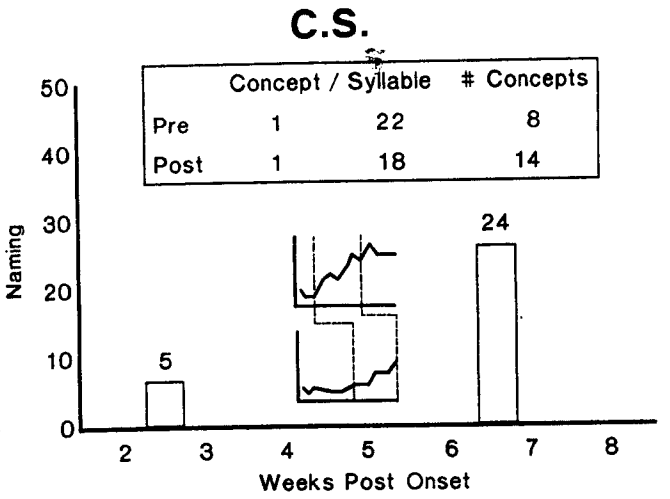

Figure 2. Performance of Subject C.S. on pre- and post-treatment responsive and confrontation naming.

Performance in pre and post testing of responsive and confrontation naming with BDAE subtests improved from 5 to 24 of a possible 48 correct (Figure 2). The quantity and efficiency of cookie-thief picture description improved from a ratio of 1 concept per 22 syllables and 8 concepts to 1 concept per 18 syllables and 14 concepts. While we cannot account for all the factors contributing to changes from pre- to post-tests, the acquisition of efficacy data on related tasks strengthens our ability to relate the influence of treatment to this patient's outcome.

Patient #2. E.L. was a 70-year-old male with a left temporal CVA who presented fluent aphasia with anomia and alexia, and an overall PICA percentile of 68. E.L. was treated for 37 sessions between 3 and 7.5 weeks post onset. The study was conducted over 13 days between 5 to 7 weeks post onset in 11 sessions.

Our goal was to improve word recognition by establishing reliable oral reading of 10 nouns and 10 verbs. The treatment utilized 3 levels through which the patient progressed if correct responses were elicited. Level 1 provided a picture, the printed word, and the patient was asked to trace letters of the word then read the word aloud. Level 2 used the same sequence but the picture was removed prior to the patient's tracing of the word. For Level 3, the patient was asked to orally read the word. If incorrect at any of the 3 levels, E.L. was asked to a) repeat the tracing, then read aloud; if not b) the clinician assisted with tracing, then E.L. read aloud; if not c) a verbal model was provided followed by patient repetition. The probe required E.L. to orally read the noun or verb. The probe was conducted prior to treatment, and lists were counterbalanced.

Scoring was plus-minus and reliability was not assessed. Treatment was to be discontinued at 80%. Pre- to post-testing included 3 BDAE subtests related to reading.

Table 2. Treatment program for Patient E.L.

CLINICAL GOAL:	Improve word recognition through oral reading
STIMULI:	10 nouns, 10 verbs
TREATMENT:	Correct Level 1 Picture-word-trace, read aloud Level 2 Picture (remove)-word-trace, read aloud Level 3 Read aloud Incorrect (within level) A. Repeat trace Read aloud B. Assist trace Read aloud C. Verbal model Patient repeat
PROBE:	Oral reading of noun/verb
SCORING:	Plus-minus
RELIABILITY:	Not assessed
CRITERION:	80%
PRE-POST:	Word-picture match, read aloud, word recognition (BDAE)

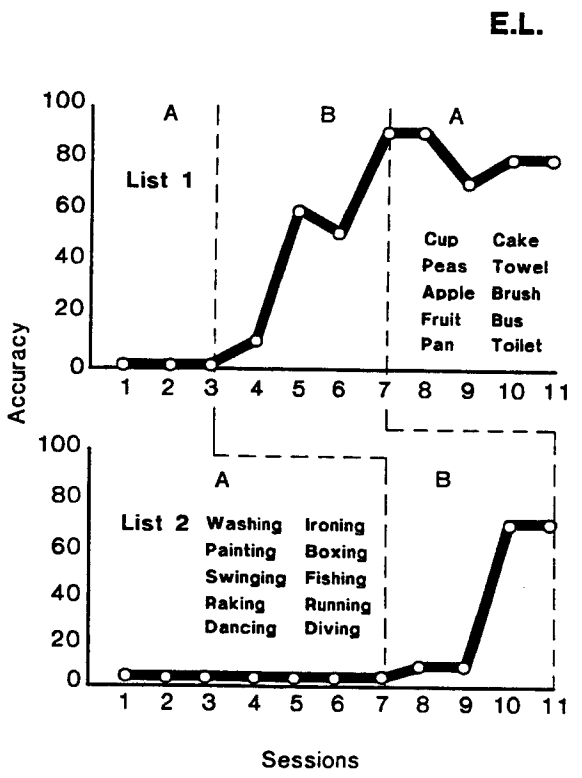

Figure 3. Performance of Subject E.L. on program to improve word recognition by means of oral reading.

A multiple baseline design showed a 0 baseline for nouns, with improvement to 90% in 4 sessions followed by maintenance averaging 82% across 5 sessions (Figure 3). Concurrently, oral reading of words remained at 0 baseline for 7 sessions, then rose to 80% in 4 sessions, at which time the patient was discharged. Pre- to post-test scores across BDAE subtests of word-to-picture matching, reading aloud and word recognition improved from 1 to 13 of 24 possible points (Figure 4). The treatment appears efficacious and parallels significant improvement in independent pre- to post-test measures.

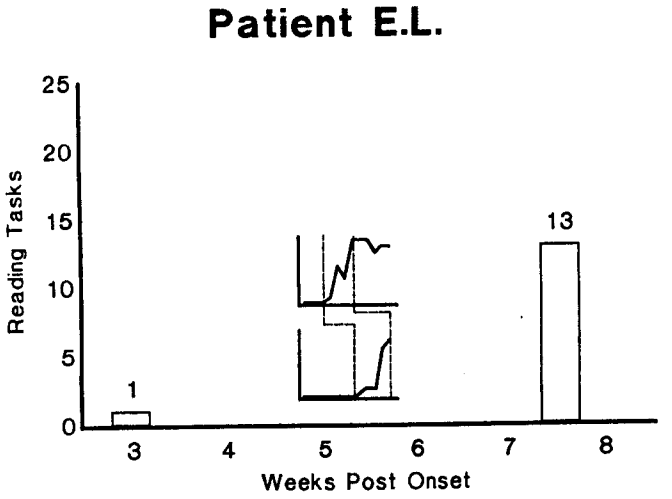

Figure 4. Performance of Subject E.L. on pre- and post-treatment BDAE reading subtests.

Patient #3. O.N. was a 71-year-old female with a left frontal parietal CVA with moderate to severe nonfluent aphasia. She was treated for 32 sessions between 4.5 and 9 weeks post onset. The study occurred over 13 days between 7 to 9 weeks post onset in 11 sessions.

The clinical goal was to improve O.N.'s word association skills by having her name objects belonging to the same semantic category as the treatment stimuli. These were 3 lists of 5 nouns each, two lists containing paired items from the same category. During treatment, a list item was verbally presented (e.g., coat) and then the category (clothing), and a second example (e.g., mittens) was presented. The patient was asked to provide a third clothing item. This always elicited a correct response from O.N. and was continued across remaining list items. Then the sequence was repeated, but without the second example. If incorrect at this level, O.N. was provided additional cueing which always elicited a correct response. Probes were conducted prior to treatment, a correct response being any item in the semantic category other than those presented in treatment. Order of lists probed was counterbalanced. Scoring was plus-minus (self-corrections allowed). Reliability was not assessed. Treatment was to be discontinued at 80%. Three BDAE subtests of naming and word fluency were selected for pre-test to post-test measurement.

Table 3. Treatment program for Patient O.N.

CLINICAL GOAL:	Improve word association
TARGET BEHAVIOR:	Name objects related to stimuli
STIMULI:	Three lists of five nouns; two lists semantically related
TREATMENT:	1) Verbal model (coat) 2) Category (clothing) and second example (mitten) 3) Patient: Provide a third item from category 4) Above repeated without second example 5) If incorrect, examples again with additional cues (e.g., body part clothed.)
PROBE:	List item presented. Patient provides another item from category.
SCORING:	Plus-minus (self-correction allowed)
RELIABILITY:	Not assessed
CRITERION:	80%
PRE-POST MEASURE:	Confrontation, responsive naming, word fluency (BDAE)

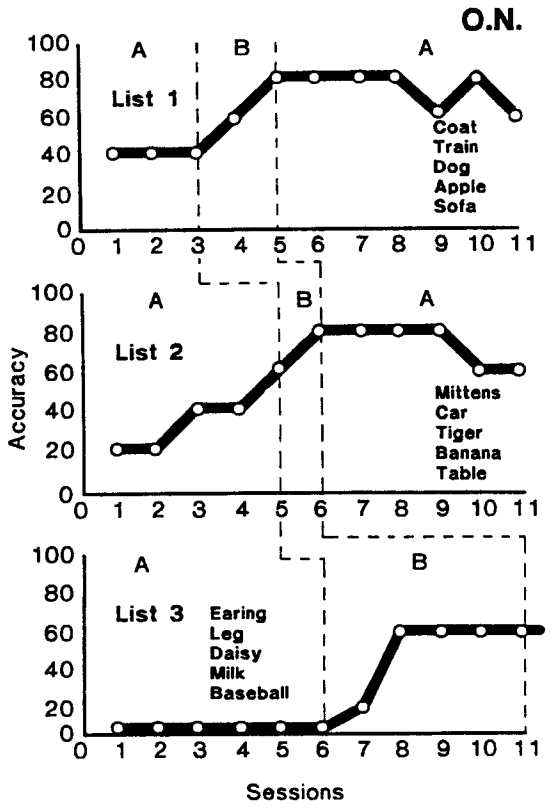

Figure 5. Performance of Subject O.N. on program to improve word association.

A multiple baseline design across the 3 sets of common nouns elicited a stable baseline for list 1 at 40%, improvement to criteria with treatment in 2 sessions and unstable maintenance over 6 additional sessions. Concurrently, list 2 (semantically related to list 1) showed an increasing trend during baseline and changed to reach criterion in a single session of treatment followed by a deteriorating maintenance trend. List 3 (not semantically related to lists 1 and 2) shows stable, low baseline across 6 sessions followed by treatment for 5 sessions, performance plateauing at the 60% level. The patient was discharged after session 11. Comparison of lists 1 and 3 provide reasonable evidence that improvement in word association for semantic categories represented by these items was due to treatment. Improvement in word association for list 2 may have been due to generalization within semantic categories or other factors.

Pre- to post-test measures of combined BDAE subtests of responsive and confrontation naming and word fluency (45 sec.) showed an increase from 14 at 4 weeks to 23 at 8 weeks post onset (Figure 6).

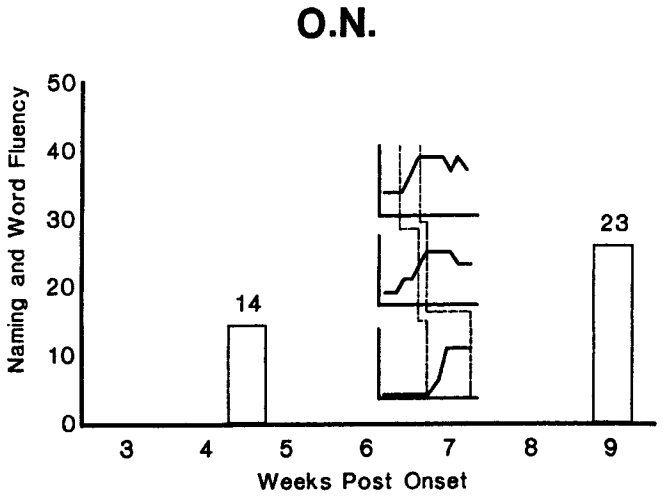

Figure 6. Performance of Subject O.N. on BDAE subtests of responsive naming, confrontation naming, and word fluency.

DISCUSSION

Constraints.

In our setting more studies fail than succeed. Restrictions in time and rapidly recovering patients make it difficult to select behaviors that are stable and treatable at baseline, apply sufficiently powerful treatments, complete necessary phases of the study, and conduct legitimate and ad-hoc changes in methodology. For instance, the evaluation of one patient indicated that object naming treated with picture and verbal description would be a suitable clinical goal and meet the requirements of a legitimate baseline behavior. By the time the baseline phase was initiated, level of performance had risen in both lists to 50% followed by an increasing baseline trend for list 2 during the treatment of list 1. Obviously naming at this level was improving on its own and did not need to be treated. A more difficult task, either within or across

modalities, should have been selected. It is worthwhile to note that, without a probe of untreated items we would not have realized our error in treating this behavior.

In another patient naming was treated using an initial phoneme cue with "wh" questions. Unfortunately, treatment was not powerful enough to elicit a selected performance level of 80%. At session 10, realizing that discharge was only several days away, we opted to initiate treatment on list 2, reaching only 60% by discharge at session 15. We could have added additional components to the treatment as in a "B-C" phase to evaluate the influence of increasing the power of our treatment, or have selected another target behavior and initiated a second study. But as is often the case, we ran out of time. We are attempting to initiate the studies earlier in the patient's length of stay, but it is often difficult to identify a stable and appropriate target behavior quickly enough.

Actually, it is relatively easy to identify a stable behavior in the acute aphasic patient. The challenge is to identify a stable behavior for which one possesses a sufficiently powerful treatment. Twenty years ago, Porch (1967) and others demonstrated that accuracy or level of performance increases as the complexity of the task changes from difficult to easy. Variability of performance tends to be greater in tasks of moderate complexity (middle of the curve and less with very difficult or very easy tasks, a notion quantified as Peak-Mean Difference (Porch, 1981). For many of our acute patients, difficult task such as writing will remain stable over time, are limited in accuracy and variability, but are not likely to be influenced by treatment. Similarly, an easier behavior such as comprehension of biographical information tends to be stable, highly accurate, and doesn't need to be treated. The medial, more variable portion of the curve may represent behaviors such as verbal labelling and reading single words.

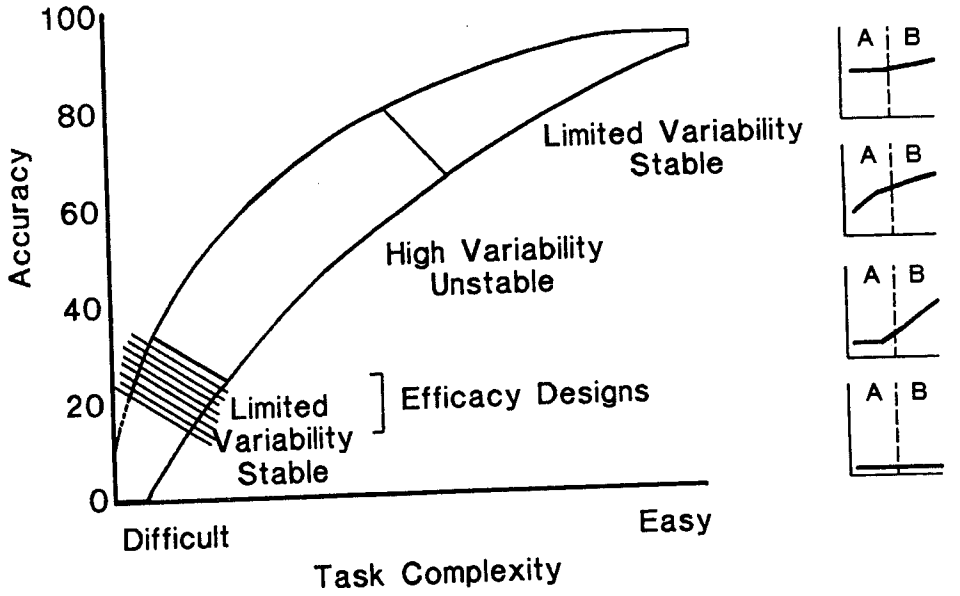

Figure 7. Task complexity and accuracy of performance and their influence on selection of baseline behaviors.

This area or "fulcrum" of the curve is where clinicians could select behaviors for treatment (after Porch, 1970). However, the same variability that renders these behaviors suitable for eliciting a desirable percentage of correct responses may produce instability or an inappropriate level of performance when probed during the baseline phase. As a result, there is a somewhat restricted "window," as we have come to call it, represented in Figure 7 by diagonal lines. This window represents the area where one can find baseline behaviors that are low in accuracy, stable in trend and variability for several days, and (given our knowledge about recovery across and within modalities) likely to be influenced by treatment. Between one day and the next, the size of the window and its location may change. Access to the window can be facilitated by changing the complexity of tasks within or across modalities and by loosening or tightening response criteria.

Costs

We have observed that the basic difference between "routine clinical procedure" and rigorous study of a target behavior is the time to probe items in nontreatment phases. Usually, one-half of a 30-minute session (15 minutes) is required to treat the behavior under study. The remainder is devoted to an average of two other treatment goals. Of that 15 minutes, 5 minutes is required to probe untreated behaviors. Currently we charge $58 per half-hour of treatment. Thus the cost of probes is roughly $10 per session. During a typical length of stay, an aphasic patient receives 35 treatment sessions for a charge of $2,030, of which $348 represents the experimental study of one target behavior over an average of 12 sessions. The total costs for probes averages $120. This represents an allocation of 6% of total treatment charges for probes, with 33% of all charges related to the study. While we can reduce these costs by probing less frequently during nontreatment phases, we maintain that allocating a portion of treatment costs to measuring the status of untreated baseline behaviors is in the best interest of the patient and the third-party payor. In our setting, the status of untreated baselines has become a reference for selecting target behaviors within the context of the patient's functional communication needs.

Currently, third-party payors accept outcome measures which are represented by documentation of change between admission and discharge. Fortunately for those of us who are providers of services, payors continue to either (a) assume our treatments were responsible for such changes, or (b) have not become adequately pressured to call the question. We suggest that these simple but legitimate multiple baselines studies will strengthen our ability to relate the efforts and cost of our treatment to patient outcome, and in so doing justify the allocation of costs necessary to demonstrate efficacy in acute rehabilitation.

REFERENCES

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

Porch, B.E. and Callaghan, S. Making predictions about recovery: Is there HOAP? In R.H. Brookshire (Ed.), Clinical Aphasiology: Conference Proceedings, 1981. Minneapolis, MN: BRK Publishers, 1981.

Porch, B.E. PICA Treatment and Prognosis in Aphasia. PICA Workshop Video-Tape Series, Alburquerque, New Mexico, 1970.

C: One of the strikes against bringing science into the clinic has always been that there's no time for it, that it costs, and that it's a penalty to the patient and to the third-party payor. I have one concern. I think it's very important that we distinguish between accountability and efficacy. One of the issues that has been debated by Spradlin, Siegel and others is that it's inappropriate to use single case designs and imply that we have valid efficacy data.

R: Our interest in conducting these studies is two-fold. One would be to try and demonstrate efficacy of our treatment in a for-profit acute setting. The other is long range -- to take advantage of efficacy of treatment as a marketing tool or as a tool for developing standards or criteria for accountability -- whether for program evaluation and quality assurance for accrediting agencies, or for third-party payors. Efficacy and accountability are two different things. I'm not sure where you draw the line between the two. Certainly these studies represent an attempt to conduct legitimate, simple studies that will fit into the constraints of a clinical operation.

C: The point is that there doesn't have to be a distinction between scientific endeavors and clinical endeavors, but my caution is this. When there are other things going on, such as other treatments and when we're collecting data on a small range of behaviors and not presenting data on those other behaviors and those other treatments, then I think we have to distinguish between efficacy and accountability. I think you've demonstrated that you have accountable data. I'd be a little more cautious about saying that you have demonstrated the efficacy of your treatment procedures. To me that's a critical distinction. I'm not sure we can do both in a realistic clinical situation and I think we avoid criticism when we make that distinction. An additional comment -- most single case designs have been with chronic patients. I've always been bothered by that. I'm not sure we know enough about the acute patient, and I think you've demonstrated that we can study those folks and learn something about them.

C: If I were paying for this and I saw that you have efficacy data on one-third of the time you spent with your patient, 12 out of 35 sessions, and that I'd gotten my money's worth for 6% of the total cost, what did you give me for the rest of the dollars?

R: You have to start with single behaviors, and I don't think third-party payors will react that way. The third-party payors are looking for patient outcome. Patient outcome is represented by two things -- what the patient does and what the medical record shows. They're accepting outcome measures that reflect change, not why the change occurred. I believe that third-party payors would be happy to have at least one more element of information that justifies or explains the need for the cost.

C: I'd like to change the topic and talk about the idea of variability with respect to task difficulty. I'd like to suggest that task difficulty and stability will interact with the severity of the aphasia.

R: Certainly these will vary with severity across patients and then will change in time as severity changes. What you have to do is manipulate a variety of things in your treatment, both in terms of the stimuli and the response, to move that window to a point where it could contain a legitimate baseline behavior. We mentioned changing the complexity of the task

either within or across modalities, or making response criteria more difficult or less difficult. These patients are two to four weeks post-onset -- they're moving pretty quickly. Half the time, by the time we decide what's a legitimate behavior, we start the baseline and the study is over. So it moves very quickly.

C: I'm wondering what was the goal of those two treatment programs. It seems to me that you had this multiple baseline across lists of items or word lists, and that you were teaching your patients to perform on these items.

R: Yes, that was the goal. The issue you might be getting to is are we teaching tasks or looking for generalization? It appears to be somewhat of a tradeoff because scientifically you want to show stability of an untreated baseline but clinically you're hoping that what you teach the patient will generalize to untreated behaviors. We've grappled with this to some extent. We're finding that the influences of treatment are fairly stimulus specific. If untreated baselines are moving we don't know if it's spontaneous recovery, generalization, or something else. So, I don't think that's our goal clinically. We want the patient's verbal labelling, for instance, to improve. It's not just those particular items that we're interested in. If stimuli are within a semantic category and constitute a response "class," you might get movement on an untreated baseline. But is that movement because those words are spontaneously getting better because of their semantic relationship, or are you getting generalization? We don't know.