

# Clinical Research Methodology in Evaluating the Therapeutic Process

D. R. VAN DEMARK, Ph.D.

*Iowa City, Iowa 52240*

Clinical speech pathologists are becoming increasingly aware of the need to define the therapy process and document more carefully the behavioral changes which occur during therapy or as a result of it. The recent work of Shelton and his associates (2, 3, 7) has provided information concerning the influence of therapy in modifying behavior, the effects of general and specific reward responses on articulation behavior, and the difficulties involved in evaluating articulation improvement by means of articulation tests.

In the area of cleft palate most of the research has dealt with the diagnosis and description of the speech of individuals with clefts, and the therapy process for these individuals has received little systematic attention. Accordingly, a study was designed in order to: (a) look critically at several *methods* of recording and measuring behavior changes during and/or as the result of therapy; and (b) apply these methods in evaluating the progress of a group of individuals actually enrolled in a therapy program. This paper is concerned with the first objective, specifically, the evaluation of: (a) a battery of tests administered before and after therapy; and (b) a method of daily observation of levels of sound production by each subject.

## Background Information

The Department of Speech Pathology and Audiology, University of Iowa, has had for many years a Summer Residential Speech and Hearing Program. Children eight years of age or older are selected from throughout the state for an intensive six-week therapy program. These children are housed in a common dormitory with supervision, recreation and cultural opportunities.

Eleven children with cleft of the palate or cleft of the lip and palate were selected for this pilot study. Their ages ranged from eight to fifteen years and none of them had participated in the program previously. Some of the children had had speech therapy in public schools or clinics or on a private basis, while others had never had therapy. All subjects had been previously seen by a speech pathologist who, in his judgment, felt that the subjects selected should make behavior changes during an intensive therapy program. The particular evaluational

procedure was determined by the speech pathologists who evaluated the subjects.

### General Procedure

Each subject was seen daily for two, one-half hour therapy sessions on an individual basis. He was also seen in a group of not more than four children for one half-hour session and for one hour session daily. Each subject therefore received two and one-half hours of therapy each day, five days a week. Because of periods of initial and final testing, holidays, and illness, the actual number of therapy days approximated twenty-five.

The clinicians were graduate students in speech pathology-audiology, all had previous experience in a practicum area—but not necessarily with children having clefts of the palate. The experimenter supervised all clinicians in their therapy programs.

No clinician worked with only one child, nor did any child work with less than two clinicians. No restrictions were placed on the number of speech sounds included in the remedial therapy program; for example, a child may have had three clinicians and may have worked on three different sounds, if the supervisor thought it advisable.

PRE AND POST-THERAPY TESTING PROCEDURE. Each subject was seen by the experimenter and his assistants during the first two days of the therapy program. During this time, the following data were obtained.

1. A tape-recorded sample of connected speech to be used in rating articulation defectiveness and nasality.
2. A tape-recorded 105 item articulation test which included the Templin-Darley Screening Test of Articulation (11), and the (IPAT) Iowa Pressure Articulation Test (6).
3. The repetition after the experimenter (12) of 13 tape-recorded sentences.
4. A discrimination evaluation of errors noted on the IPAT.
5. A tape-recorded stimulation evaluation on all errors noted on the IPAT.
6. A hearing evaluation.
7. Manometer ratios (5).
8. A description of the oral mechanism as well as lateral X-rays and/or cinefluorographic films.
9. A Peabody Picture Vocabulary Test.

At the conclusion of therapy, the above tests were re-administered with the exception of discrimination, hearing, and Peabody Picture Vocabulary Tests.

PLAN OF THERAPY. All individual and all half-hour group sessions were formal therapy sessions. The hour group sessions were activity-oriented-speech therapy but such activities were planned to constitute no more than 50% of the hour. These sessions were provided to help children learn

to use newly learned sounds in a controlled, but less formal environment. Clinicians, in turn, were able to observe changes, or lack of change, in behavior in the less structured situations.

To aid in communication among clinicians, each subject had a notebook of assignments which was available during each therapy session. Each clinician was responsible, with the guidance of the experimenter, for the planning of the therapy program of one subject. Therefore, each clinician was in charge of coordinating therapy so that a given child would not become confused in working with more than one clinician.

Clinicians were required to write logs for each therapy session which described the goals of the therapy, the methods used, the sound or sounds emphasized, the average level of each subject's sound production, and the amount of therapy time actually spent in specific drill. Twenty-seven arbitrarily defined levels of sound production and discrimination ability were used to rate the subject's average level of learning for a given therapy session. These levels of production learning were determined for each child in all therapy sessions. The levels of learning are noted in Figure 1. Different symbols were used to denote whether the subject was in an individual or group therapy session. In the example presented in Figure 1, three clinicians made separate ratings of the level of learning of the /s/ sound for a given subject.

Furthermore, tape recordings were made weekly of each subject's speech. A sample of conversational speech was elicited and each subject repeated sentences containing sounds that had been emphasized in therapy during the week. It had also been planned to have an observer evaluate each

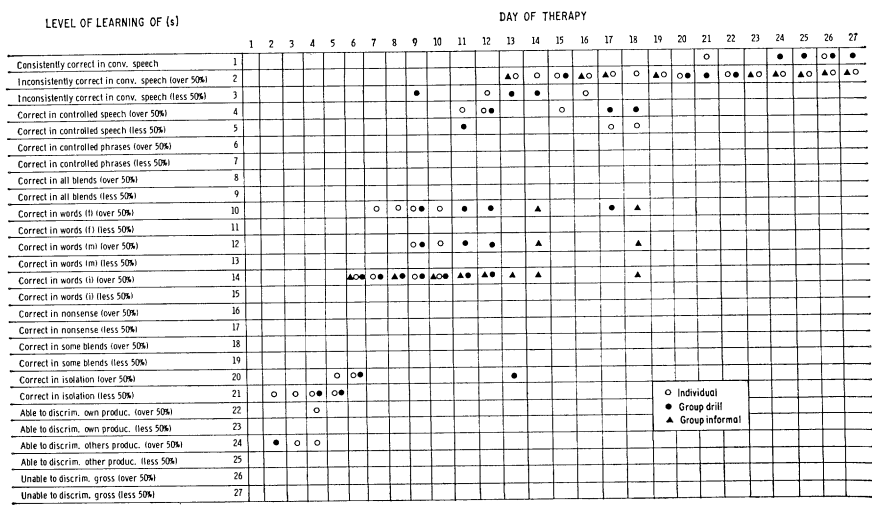


FIGURE 1. Level of learning of /s/ for a given subject who received therapy on this sound by three different clinicians. Day of therapy and type of therapy sessions are noted.

subject's level of sound production at weekly intervals for all therapy sessions for reliability purposes. However, since observation rooms were not available at the time of this experiment, this part of the experimental program was deleted. The reliability of clinician reporting, however, can be observed in Figure 1 by comparing ratings by different clinicians who were working with a subject on the same sound.

### Data Evaluation

At the end of the experimental program the data obtained for each measure were examined. It appeared that each measure used offered some advantages and disadvantages, which are described below.

**JUDGED ARTICULATION DEFECTIVENESS AND NASALITY.** Judgments of articulation defectiveness and nasality, obtained by having judges rate samples of conversational speech on a seven point equal-appearing intervals scale, are valuable tools for listeners' perceptual rating of subjects. This type of rating can be done reliably by a group of judges (8), but not by an individual clinician (1).

The use of these ratings in rank ordering subjects is appropriate in many instances; yet these ratings are not descriptive or discriminating and may not be used to evaluate changes during and/or after therapy. For example, if a subject receives a pre-therapy rating of articulation defectiveness of *five*, moderately severe, and receives a post-therapy rating of *four*, moderate, "improvement" is implied. However, such information gives no description of the improvement that took place or the way in which it occurred.

**ARTICULATION TESTS.** Pre- and post-therapy scores on the 105 item articulation test, the Templin-Darley Screening Articulation Test (11), and the Iowa Pressure Articulation Test (IPAT) (6) can be compared. Changes in scores should indicate "progress" and, by examination of the test item, specific changes can be identified. Several factors, however, should be considered when making such comparisons. For example, if a subject omitted the /r/ and /s/ sounds at the beginning of therapy and if all other sounds were correct, he would miss 24 items on the Templin-Darley and 23 items on the IPAT. If therapy were initiated and the subject improved these sounds to mild distortions, no change in score would be evident. On the Templin-Darley Screening Test and on the IPAT, there are many blends consisting of fricatives, plosives, and glides. A child may be able to make many of these sounds as single items, yet fail to do so in blends (Spriestersbach et al., (10), Morris et al. (6)) and thus he achieves a score which indicates little or no behavior modification. Although these test scores may indicate progress on particular sounds if a subject learns to produce them adequately, no systematic approach is available to demonstrate degree of progress. As has also been demonstrated previously, children are inconsistent in their misarticulations (Spriestersbach et al., (9)) and a sample from one word may not be

representative of the way the child articulates the sound in other words, nor is it logical to assume that a sound which is correct on a word articulation test is also correct in conversational speech. The Templin-Darley Screening Test and IPAT serve their purpose of screening and diagnosis but fail to provide adequate methods of rating progress.

**REPEATED SENTENCES TEST.** Thirteen sentences repeated after the experimenter (12) can yield measures of articulation proficiency for total sounds, for manner of production categories, and for specific sounds. In these 13 sentences, consonants are tested repeatedly with a frequency approximately proportional to their frequency of occurrence in the English language. For example, the phoneme /d/ is tested 12 times. It is therefore possible to derive a consistency score for each sound tested; i.e., if /d/ is produced correctly nine times out of twelve, the consistency score for that sound is 75%. Figure 2 demonstrates the use of this score in measuring the progress of an individual subject. Plotting profiles for pre and post-therapy administrations of the test reveals changes in the consistency of correct articulation of several sounds.

Using the same repeated sentences it is also possible to display graphically the changes in the number of misarticulations according to manner-of-production categories (i.e. plosives, fricatives, etc.) Figure 3 demonstrates the manner-of-production profiles of two subjects.

The repeated-sentences task has two disadvantages. Scoring the sample requires some sophistication, although it can be done reliably (12). Also, because of the controlled frequency of occurrence of phonemes, a few

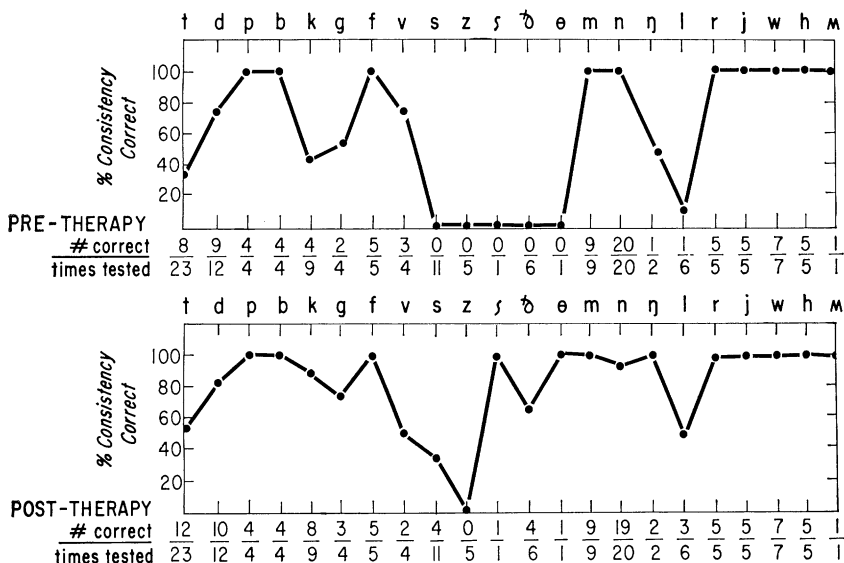


FIGURE 2. Pre- and post-therapy scores for a subject for each phoneme tested as measured by the sentence consistency test. Per cent of consistency of correct production is plotted.

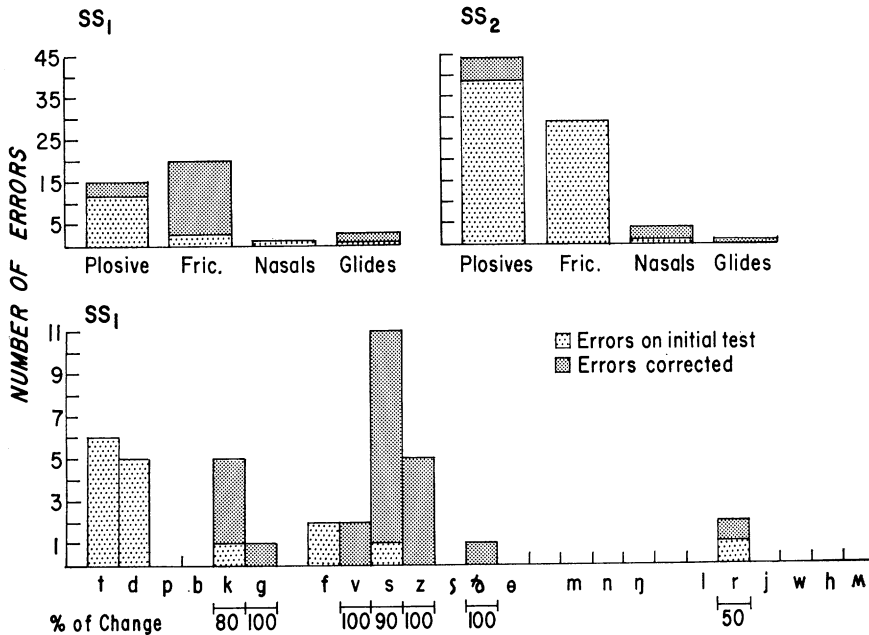


FIGURE 3. Comparison of behavior changes made for two subjects when number of errors on manner of production categories is considered. Degree of specific error modification is also indicated for SS<sub>1</sub>.

relatively-infrequently-occurring consonants are tested only once. The task has certain advantages over typical word-articulation tests in that (a) it provides a controlled sample of speech which has a demonstrated correlation of .93 with spontaneous connected speech (12); (b) sounds which occur frequently in the English language are tested repeatedly, and (c) individual scores can be plotted on profiles, making it easier to document and visualize progress.

**DISCRIMINATION TESTS.** In pre-therapy data collection, each subject was required to listen to words which contained his identified error sounds. For example, if a subject failed to produce an adequate /k/ during articulation testing, he was asked to listen to repetitions of a word containing that sound on a language master card. These words were presented in random order with the sound omitted, distorted orally and/or nasally, substituted, and correct. Subjects had little difficulty with this task and it was concluded that errors made in discriminating the "correct sound" were minimal and could, in fact, be attributed to chance. Discrimination tests, therefore, were not readministered at the conclusion of the therapy.

Observations made during therapy would suggest that a subject may fail to discriminate errors in his own speech, but has little difficulty discriminating errors in other contexts; i.e., the articulation of others.

**STIMULATION TESTING.** Following the procedure suggested by Milisen (4) stimulation testing was executed for all error sounds noted on the 43

item IPAT. Originally, it had been planned to test for all error items on the 105 item articulation test; but subjects had a great many errors, and time and subjects' attention spans did not allow for this type of thorough testing procedure. All testing was recorded on a high quality tape recorder. Each test item missed on the IPAT was retested, using another picture. The experimenter then repeated the desired stimulus twice with the subjects responding immediately. Stimulation was provided for sounds in isolation, nonsense syllables (initial, medial and final positions), and words. The subject was then asked to respond to a picture containing the stimulus sound. The subject was made aware that this word contained the stimulated sound and he attempted to produce the word correctly. The subject next read a word with the stimulus sound and then read a sentence with the stimulus sound.

Examination of the data revealed several problems in this type of testing. First, subjects were very inconsistent in articulation errors: some sounds which were in error in administration of the first IPAT test were correctly produced on the second IPAT test. (Therefore, it is logical to assume that on further testing those sounds evaluated as correct initially might also have been in error). Secondly, subjects were inconsistent in difficulty producing the sound in isolation, yet would approach correct production on a stimulated word. At this stage of sophistication, it appears difficult to use stimulation testing as a research tool; however, stimulation testing combined with consistency measures appears most beneficial in designing a logical therapy program, and for this reason such testing merits further consideration in clinical research programs.

**DAILY THERAPY PLOTTING.** In addition to pre and post-therapy measures which can be compared to evaluate progress, data were also collected on a day-to-day basis. Each subject's level of performance was evaluated for each therapy session. Figure 1 demonstrates the method of plotting sound production for an individual subject. For example, when therapy started the subject was able to discriminate a correct /s/ production in the speech of the clinician over 50% of the time (Level 24). The average level of sound production attained in each therapy session was recorded, as well as whether the session was individual or one of the two group sessions. As is demonstrated, this subject's level of sound production improved from Level 24 to Level 1. It can be noted that on day 13 the therapy session at Level 20 was in all probability inappropriate since, on the same day, a clinician reported that the subject inconsistently, was able to produce the sound correctly in conversational speech (Level 2). In all probability, the spread of the levels on this day indicated at least that: (a) the child was making very rapid changes in behavior on that particular day; or that, (b) the therapy sessions were poorly coordinated with the specific goals undefined. It is of interest to note that on following days this subject was able to continue to produce the sound at Levels 2 and 3, thus indicating a more stabilized production of the sound at a higher level of behavior modification.

This chart (Figure 1) is extremely important in the evaluation of therapy procedures of clinicians. It very quickly enables a supervisor to cross check the appropriateness of an individual therapy session and a particular clinician's coordination of therapy. Another gratifying aspect of this chart is that it enables a clinician to visualize progress or lack of progress for a particular child. Although, in the experiment described here, neither the children nor the clinicians were shown the charts it appears that doing so at least at various stages in the therapeutic process would be beneficial for motivational purposes. After using this chart, it is obvious to us that more descriptive categories should be included, especially at the assumed higher levels of behavioral change. It might also be of interest to examine a wide range of behavioral level, such as highest level of performance and lowest level of performance.

Although clinicians may feel that they do not have time for writing logs, certainly some estimate must be made of a subject's or client's level of sound production if appropriate work is to begin at the next therapy session. Often in the writing of logs, much description is used and yet little is said, (for example: "The child did well on repetition of words."). One should report whether the child was able to produce the sound correctly all the time, with all blends or some blends and in all positions. It is suggested that perhaps such a plotting of progress is beneficial to the clinician in observing behavior changes and in noting whether various therapeutic methods are beneficial to change.

It was also evident that a method of observing over-all change could be noted by the level at which the child began work and the levels he attained at the termination of therapy. Although these levels are not in equal-appearing-intervals, either in achievement or difficulty, they present an easy estimate of the days required for behavior changes. Table 1 demonstrates for one subject that 26 therapy days were spent to help the child progress from Level 24 to Level 1 for the /s/ sound. The same child was able to start the production of /z/ at Level 4 and reached Level 1 in 13 therapy days. It can also be noted that consistency scores reveal that both sounds were 100% in error at the beginning of therapy with the /z/

TABLE 1. An example of sounds included in an individual subject's therapy program with the level of average behavior exhibited at the beginning of therapy, days of therapy, and level of behavior when therapy was terminated.

<i>sound</i>	<i>level at which therapy began</i>	<i>days of therapy</i>	<i>level at which therapy terminated</i>
s	24	26	1
θ	20	21	1
ð	14	19	1
z	4	13	1
tʃ	20	9	2
d	20	3	10



sound consistently correct and the /s/ sound 91% consistently correct on post-therapy testing.

**WEEKLY RECORDINGS AND OBSERVATIONS.** An attempt was made to record each subject's production of sentences containing those sounds included in the therapy program as well as a sample of connected speech at weekly intervals. However, the procedure used in this study proved inadequate. The recordings and controls were poor for this particular task; and inspection of the data indicates that it is unreliable. Since no observational facilities were available at the time of the experiment, it was also necessary to delete the planned observation periods. It would appear, however, that both recordings and observations at weekly intervals were important in the substantiating clinical reliability and both methods should be used in future investigations.

## **Discussion**

This type of experimental program appeared to have no adverse effect on subjects when compared to other children involved in previous therapy programs, or to children who were not involved in the research program. Clinicians were cooperative and reported that the program aided in structuring the therapeutic situation. Although the experimenter felt that the requirement of noting the amount of time spent in therapy on a particular sound was cumbersome, clinicians reported that this requirement helped them use their time more efficiently.

It can be questioned whether the above therapy program offers sufficient opportunity for clinical experimentation and trial and error, which can be considered important in the student-clinician learning experience. This type of experimental therapy program might be more appropriately executed with professional clinicians; however, the methods and backgrounds of clinicians might still vary greatly.

At this stage of clinical research many questions concerning the above methodology can be asked and few answers given. It is our feeling that the consistency scores and the plotting of daily levels of sound production appear most promising as clinical research tools. It is felt now that few of the presently used testing procedures can be discarded, but, certainly more appropriate testing procedures need to be developed.

## **Summary**

Eleven subjects with clefts of the palate were submitted to an intensive therapy program for a period of six weeks. Various measures, including articulation tests, stimulation tests, consistency tests, and plotting of daily progress, were obtained before therapy, during therapy, and after therapy. The usefulness of these measures in defining changes in articulatory behavior during the therapeutic process is discussed. It appears that consistency scores and plotting of daily behavior changes are promising tools in defining the therapeutic process. It is obvious that much research is

needed in determining methods for measuring therapeutic change and in describing the therapeutic process.

*Acknowledgment:* This report is based on research which was supported in part by PHS Research Grant DE 00853, National Institute of Dental Research.

## References

1. BRADFORD, L. F., BROOKS, ALTA R., and SHELTON, R. L., JR., Clinical judgment of hypernasality in cleft palate children. *Cleft Palate J.*, 1, 329-335, 1964.
2. CHISUM, LINDA, SHELTON, R. E., JR., ARNDT, W. B., JR., and ELBERT, MARY, The relationship between remedial speech instruction activities and articulation change. *Cleft Palate J.*, 6, 57-64, 1969.
3. ELBERT, MARY, SHELTON, R. L., JR., and ARNDT, W. B., JR., A task for evaluation of articulation change: I. Development of methodology. *J. Speech Hearing Res.*, 10, 281-288, 1967.
4. MILISEN, R. L., and ASSOCIATES. The disorders of articulation: A systematic clinical and experimental approach. *J. Speech Hearing Dis.* Monograph supplement, 4, 1954.
5. MORRIS, H. L., The oral manometer as a diagnostic tool in clinical speech pathology. *J. Speech Hearing Dis.*, 31, 362-369, 1966.
6. MORRIS, H. L., PRIESTERSBACH, D. C., and DARLEY, F. L., An articulation test for assessing competency of velopharyngeal closure. *J. Speech Hearing Res.*, 4, 48-55, 1961.
7. SHELTON, R. L., JR., ELBERT, MARY, and ARNDT, W. B., JR., A task for evaluation of articulation change: II. Comparison of test scores during baseline and lesson series testing. *J. Speech Hearing Res.*, 10, 578-585, 1967.
8. SHERMAN, DOROTHY, The merits of backward playing of connected speech in the scaling of voice quality disorders. *J. Speech Hearing Dis.*, 19, 312-321, 1954.
9. PRIESTERSBACH, D. C., DARLEY, F. L., and ROUSE, VERA, Articulation of a group of children with cleft lips and palates. *J. Speech Hearing Dis.*, 21, 436-445, 1956.
10. PRIESTERSBACH, D. C., MOLL, K. L., and MORRIS, H. L., Subject classification and articulation of speakers with cleft palates. *J. Speech Hearing Res.*, 4, 362-372, 1961.
11. TEMPLIN, MILDRED C., and DARLEY, F. L., *The Templin-Darley Tests of Articulation*. Bureau of Educational Research and Service, Extension Division, University of Iowa, Iowa City, Iowa, 1960.
12. VAN DEMARK, D. R., Misarticulations and listener judgments of the speech of individuals with cleft palate. *Cleft Palate J.*, 1, 232-245, 1964.