

Articulatory Changes in the Therapeutic Process

D. R. VAN DEMARK, Ph.D

Iowa City, Iowa 52240

In a previous report the methodology employed to study articulatory changes in a group of 11 children with clefts of the palate has been reported (8). In the initial report the chief concern was to evaluate measures which can be used to identify behavioral changes. The advantages and disadvantages of the measures employed were discussed and defined and the present article will be more meaningful if the reader is familiar with this report. The question asked in this report was: What measures demonstrate significant differences in pre and post therapy scores?

Procedure

Subjects received two, one-half hours of individual therapy and one and one-half hours of group therapy daily for a period of six weeks. Generally, therapy involved procedures and methods typically used with children having functional articulation problems. Basic emphasis throughout therapy was placed on improving articulation.

Hearing evaluation indicated that no subject exhibited a hearing loss of greater than 15 dB in the better ear for the speech frequencies (500, 1000, 2000 Hz., ASA standards). Six of the 11 subjects exhibited a conductive hearing loss in one ear and ten of the 11 subjects were active patients at University Hospitals, Department of Otolaryngology.

Subjects averaged a Peabody Picture Vocabulary Test vocabulary recognition intelligence quotient of 105 with a range from 68 to 128. The subject with the lowest score was bilingual and standard intelligence tests had placed this child and all other subjects above the mentally deficient range.

Each subject was seen by the experimenter and his assistants during the first two days of the therapy program. During this time the following tests and data were obtained:

1. A tape recorded sample of connected speech to be used in rating articulation defectiveness and nasality.
2. Tape-recorded 105 item articulation test which included the Templin-Darley Screening Test of Articulation (6) and the (IPAT) Iowa Pressure Articulation Test (4).
3. Repetition of 13 tape-recorded sentences after the experimenter (7).

4. A discrimination test on errors noted on the IPAT.
5. A tape-recorded stimulation test on all errors noted on the IPAT.
6. Hearing Tests.
7. Manometer ratios (3).
8. A description of the oral mechanism as well as lateral x-rays and/or cinefluorography.
9. A Peabody Picture Vocabulary Test.

At the conclusion of therapy, the above battery of tests were readministered, with the exception of discrimination, hearing, and the Peabody Picture Vocabulary Tests.

The reliability of listener judgments in scoring the various articulatory measures was determined by computing the percentage of agreement for two listeners. Listener agreement on the various articulation tests ranged from 91% to 94% and reliability was considered sufficient for the purposes of this study.

Results

Severity Ratings. Thirteen judges rated a sample of conversational speech obtained from each subject in both the pre and post-therapy conditions. Using a seven point equal-appearing-intervals scale, the average rating for severity of articulation defectiveness was 4.72 for the pre-therapy condition and 3.64 for the post-therapy condition. The difference was significant at the 5% level of confidence ($t = 2.43$). When the conversational speech samples were rated for severity of nasality by use of a seven-point scale, there was a significant difference ($t = 2.10$) between the pre and post-therapy conditions. Judges also rated severity of articulation defectiveness on the speech samples which were comprised of thirteen repeated sentences. The difference between the pre-therapy ratings (4.15) and post-therapy ratings (3.54) was significant at the .01 level of confidence ($t = 3.23$). The reliability for averaged ratings ranged from .97 to .98 (2).

To determine whether the repeated sentence samples were representative of conversational speech samples, mean scale values obtained on the two types of samples were compared. The obtained correlation coefficient, computed between the two sets of values, was .93.

Articulation Scores. a) Iowa Pressure Articulation Test. On the 43-item Pressure Articulation Test only one subject failed to show improvement between the pre and post-therapy conditions. A t test revealed that the difference in articulation scores was significant at the 1% level of confidence ($t = 6.50$). It is somewhat surprising to find that the difference in scores was this great (average improvement of seven items) as the IPAT is composed of sounds which discriminate between the articulatory performance of children who achieve velopharyngeal closure and those who do not. As this test includes many pressure-blend sound combinations one would not necessarily assume that progress would be so apparent on this

test, although some sounds which were emphasized in therapy are included in the test. It would seem that when subjects fail to make progress on this measure the adequacy of velopharyngeal function should be examined critically.

b) 105-Item Articulation Test. This test consisted of the Templin-Darley Screening Test of Articulation and the Iowa Pressure Articulation Test as well as single consonant items in the initial, medial and final positions. Pre and post-therapy scores were significantly different at the 1% level of confidence ($t = 4.55$). The mean number of items corrected was 13 and only one subject failed to show a positive change in the post-therapy score.

Since this is a more thorough diagnostic measure, one would expect subjects to show more changes on this test than on the other articulatory measures administered. It may also be logical since there are more items which are not blends that if correct sound productions are learned they may be more evident on single items than on blends and thereby improve the articulation score on this test.

c) Templin-Darley Screening Test of Articulation. Comparison of pre and post-therapy scores indicated that although scores generally improved, the difference was not significant ($t = 1.85$). Two subjects achieved lower scores on the post-therapy evaluation, while the remaining subjects demonstrated varying degrees of improvement. The manner in which this test is scored (as are most articulation tests) may account in part for the apparent lack of change in articulation skills pre and post-therapy since items are scored correct/incorrect and do not reflect sound production which falls short of "correct". For example, a change from omission to mild distortion continues to be scored "incorrect" even though clinically such a change may reflect significant progress.

Repeated Sentence Test. Total number of errors and manner of production errors were compared for the pre and post-therapy recording. Significant differences were evident at the 5% level of confidence or greater for the total number of errors ($t = 3.80$), as well as errors on plosive ($t = 2.75$) and fricative ($t = 3.80$) sounds. Although the difference in number of errors on glides was not significant, there were fewer errors in the post-therapy condition. Of the 149 sounds evaluated on this measure the total number of errors on the pre-therapy condition averaged 49 with an average change in score of 12 in the post therapy condition.

Discrimination Scores. In pre-therapy data collection, each subject was required to listen to words which contained his scored error sounds. For example, if a subject failed to produce an adequate /k/ on articulation testing, he was asked to listen to repetitions of a word containing that sound on a language master card. Words were presented in random order with the sound omitted, distorted orally and/or nasally, substituted and correct. The subject was required to identify the correct word production. Subjects exhibited little difficulty with this task. The mean correct identification of error sounds was 91% with a range of 74–100%. Since subjects

were able to identify errors in the speech of others to such an accurate degree, discrimination tests were not administered at the conclusion of therapy. It would appear that if discrimination difficulties occur, they involve the subject's discrimination of his own speech rather than the speech of others.

Stimulation Scores. For all errors which occurred, both on initial and final administration of the IPAT, subjects were given a stimulation test. For all sounds scored as errors on the IPAT, subjects responded to another picture of the test sound. After the subject's response, he was then instructed to respond after stimulation of the sound in isolation, nonsense syllables, words and reading of the sound in a word and in a sentence. Of the 314 test sounds which were in error on the initial test, 12 were produced correctly on the second attempt. Of the 253 sounds in error on post-therapy examination 14 were produced correctly on the second attempt.

Two different methods of approach were used to evaluate the results of stimulation testing. Method I consisted of examining the results of stimulation testing and arbitrarily placing scores in three categories: a) correct, in at least one stimulation category, e.g. in isolation, b) change, either positive or negative during the stimulation of that sound, c) no change on any of the stimulation. Table 1 presents the distribution of scores in the above categories.

Before therapy was initiated 66% (314) of the sounds tested on the IPAT were in error. Upon stimulation testing, subjects were able to produce correctly 22% of the error sounds, change 71%, with the remaining 6% exhibiting no change. After therapy, 53% of the sounds tested on the IPAT were in error. Stimulation testing indicated that 30% of the error sounds could be produced correctly, 60% changed, and 10% did not change.

It was obvious that subjects were generally able to modify behavior, as 90% of the responses fell within the correct or change categories. Examination of the data indicated wide variation in performance as some subjects were able to modify behavior although on previous testing they had been unable to do so.

Method II was devised to examine the "change" category more thoroughly and to attempt to evaluate performance at various levels of difficulty. Each initial error response was considered as a base score. Responses were ranked from least to most severe as follows: correct, mild oral-

TABLE 1. Stimulation scores obtained by Method I. Correct indicates correct production of a sound in one or more of the eight categories, change + or - indicates a change following stimulation from the initial error, no change indicates that stimulation did not affect subjects production in any situation.

<i>Number of errors</i>		<i>Correct</i>	<i>Change (+ or -)</i>	<i>No Change</i>
Pre-therapy	314.....	70	224	20
Post-therapy	253.....	76	152	25

TABLE 2. Stimulation scores by Method II. N refers to number of sounds in each stimulation category which were either correct, changed in a positive direction, changed in a negative direction, or exhibited no change from the baseline score.

	<i>Isolation</i>		<i>Nonsense Syllables</i>						<i>Word</i>		<i>Controlled Word</i>		<i>Reading Word</i>		<i>Sentence</i>	
			<i>I</i>		<i>M</i>		<i>F</i>									
	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Correct Responses																
Pre	33	11	15	5	12	4	15	5	24	8	25	8	18	6	13	4
Post	31	12	10	4	15	6	30	12	26	10	34	13	30	12	23	9
Positive Change																
Pre	96	31	54	17	84	27	71	23	104	33	106	34	96	31	100	32
Post	78	31	56	22	57	23	58	23	51	20	57	23	53	21	59	23
Negative Change																
Pre	46	15	33	11	61	19	48	15	89	27	85	27	78	25	103	33
Post	32	13	74	29	56	22	49	19	73	29	68	27	72	28	72	28
No Change																
Pre	139	44	212	68	157	50	180	57	97	31	98	31	122	39	98	31
Post	112	44	113	47	125	49	116	46	103	41	94	37	98	39	99	39

distortion, moderate oral-distortion, severe oral-distortion, mild nasal-distortion, moderate nasal-distortion, severe nasal-distortion, substitution-oral, substitution-nasal, substitution-glottal stop, substitution-pharyngeal fricative and omission. Performance on each of the eight levels of stimulation (see Table 2) was evaluated as: correct, positive change, negative change, or no change in relation to the base score. Table 2 indicates the scores obtained when stimulation results were summarized in this manner.

According to the results of a study by Scott and Milisen (5) one would expect that children with functional articulation errors should be able to perform best at level one (isolation), with a decrease in performance level—as level of difficulty presumably increases. When considering correct productions only, subjects in this study were in fact able to perform best at level one (isolation) in the pre-therapy condition (correct production 11% of the time). Their performance in words and controlled words was correct more often than their performance in nonsense syllables and reading sentences.

When the positive change category was considered, approximately one-third of the responses were improved in the word or controlled word contexts. Improvement in all contexts but nonsense syllables occurred over 30% of the time. Negative change occurred least frequently on isolated syllables and most frequently when reading sentences.

In an effort to further examine the influence of stimulation upon behavior modification, the no change category was examined. No change was most common in the nonsense syllable context (ranging from 50 to 60%) and least common in words, controlled words and sentences (31%).

After therapy, sounds were produced correctly on stimulation most frequently in the controlled word environment (13%). Initial and medial nonsense syllables ranked low in percent of correct production. Improved production occurred most frequently in the isolated position (31%) while

improvement on the remainder of the stimulation tasks ranged from 20% to 23%. The fewest negative changes occurred on isolated syllables (13%) while negative change was exhibited on 19% to 29% of the remaining categories.

When the no change category was considered for the various stimulation tasks, subjects showed less variation than in the pre-therapy situation. Scores ranged from 37% to 49% with the most change occurring in the controlled word environment and the least change occurring in the nonsense syllables in the medial position.

Manometer Ratios. For the total group of subjects there was no change in pre and post-therapy manometer ratios. The average ratio with bleed for pre-therapy was .85 and the average post-therapy ratios were .85.

Discussion

The results of this study indicate that various articulation measures are beneficial in determining progress made in the therapeutic process by a group of individuals with clefts of the palate. Scores obtained on severity ratings, two articulation tests, and the repeated sentence test all indicate significant differences on the post therapy testing situation. These measures can be made reliably and can be presented in a standardized manner. The validity of these measures can be questioned in that it is impossible to know if, in fact, these measures are "truly representative of the progress made". The same problem occurs with the administration of any articulation test in that the responses obtained on a given test may or may not be representative of a particular subject's speech. It appears, however, that the above measures are at least a sample of those behavior changes.

Several factors must be considered in selection of measures used in this type of therapy program. Subject selection plays an important role in determining progress. For example, if a subject does not exhibit an adequate mechanism, e.g. velopharyngeal closure, his progress in therapy will not be as great as a subject who has an adequate mechanism. Likewise, the motivation of the subject, dentition, hearing, intelligence, previous therapy, clinician's abilities, and other factors all play an important role in a subject's ability to modify his behavior.

When all subjects are considered as a group, individual differences relative to the individual's performance may not be apparent. In this study, three subjects made little progress on the various articulation and repeated sentence tests, whereas, others made varying degrees of improvement. Test information, related to other measures, e.g. velopharyngeal closure may be important in providing additional treatment considerations for a subject.

Various articulation and physical measures are beneficial in determining changes in the therapeutic process. Scores on word articulation tests are easily obtainable, and may show significant improvement, yet often fail to give information about changes in conversational speech. Changes

in repeated sentence scores from pre to post-therapy conditions may be more nearly representative of behavior modifications that are evident in conversation.

Stimulation testing may yield information to clinical research which is also applicable to the clinical situation. For example, Barnes, and Morris, (1) and others have indicated that stimulation testing procedures may be of value prognostically. In this study the results of stimulation testing are not clearly meaningful in that 1) sounds were often tested as blends, 2) sounds which had previously been correct were not tested, 3) the stimulation tasks were not randomized. With these limitations in mind, the results of the present study point out a need to further study the usefulness of stimulation testing as a diagnostic, therapeutic and research tool. Certainly if subjects are able to make essentially the same changes in more complex tasks (word production vs. isolated sound production) then it seems feasible to direct attempts to work at the level attainable.

Summary

Eleven subjects with cleft of the palate were submitted to an intensive therapy program for a period of six weeks. Various measures were obtained before and after therapy. Subjects demonstrated significant improvement on general ratings of articulation defectiveness in conversation, sentences and nasality. Two articulation tests and three repeated sentence measures were significantly different on post-therapy evaluation. Stimulation scores were evaluated by two methods. By stimulation subjects were often able to modify behavior and it appears questionable that this behavior is modified to the highest degree at the lowest level of learning.

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

References

1. BARNES, IDA and H. L. MORRIS, Interrelationships among oral breath pressure ratios and articulation skills for individuals with cleft palate. *J. Speech Hearing Res.*, 10, 506-514 (1967).
2. EBEL, R. L., Estimation of the reliability of ratings. *Psychometrika*, 16, 407-424 (1951).
3. MORRIS, H. L., The oral manometer as a diagnostic tool in clinical speech pathology. *J. Speech Hearing Dis.*, 31, 362-369 (1966).
4. MORRIS, H. L., D. C. SPIESTERSBACH, and F. L. DARLEY, An articulation test for assessing competency of velopharyngeal closure. *J. Speech Hearing Res.*, 4, 48-55 (1961).
5. SCOTT, D. A., and R. L. MILISEN, The effectiveness of combined visual-auditory stimulation in improving articulation. *J. Speech Hearing Dis.*, Monograph Supplement 4, 52-56 (1954).
6. TEMPLIN, MILDRED C., and F. L. DARLEY, *The Templin-Darley Tests of Articulation*. Bureau of Educational Research and Service, Extension Division, University of Iowa, Iowa City, Iowa (1960).
7. VANDEMARK, D. R., Misarticulations and listener judgments of the speech of individuals with cleft palate. *Cleft Palate J.*, 1, 232-245 (1964).
8. VANDEMARK, D. R., Clinical research methodology in evaluating the therapeutic process. *Cleft Palate J.*, 8, 26-35 (1971).