

*MODIFICATION OF ARTICULATION ERRORS
OF PRE-DELINQUENTS BY THEIR PEERS^{1,2}*

JON S. BAILEY, GARY D. TIMBERS,³ ELERY L. PHILLIPS,
AND MONTROSE M. WOLF

UNIVERSITY OF KANSAS

Pre-delinquent peers in Achievement Place (a community based family style rehabilitation program based on a token economy) were given points (token reinforcement) to modify the articulation errors of two boys. In Experiment I, using a multiple baseline experimental design, error words involving the /l/, /r/, /th/, and /ting/ sounds were successfully treated by both a group of peers and by individual peers. Also, generalization occurred to words that were not trained. The speech correction procedure used by the peers involved a number of variables including modelling, peer approval, contingent points, and feedback. The individual role of each of these variables was not experimentally analyzed, but it was demonstrated that peers could function as speech therapists without instructions, feedback, or the presence of an adult. It was also found that payment of points to peers for detecting correct articulations produced closer agreement with the experimenter than when they were paid points for finding incorrect articulations. The results were replicated in a second experiment with another subject who had similar articulation errors. In addition, the second experiment showed that peer speech correction procedures resulted in some generalization to the correct use of target words in sentences and significant improvements on standard tests of articulation.

Delinquents and pre-delinquents are currently receiving the interest and attention of applied behavioral researchers (Burchard and Tyler, 1965; Schwitzgebel, 1964; Thorne, Tharp, and Wetzel, 1967; Tyler and Brown, 1968). These studies have demonstrated the usefulness of a variety of behavioral procedures such as timeout, behavior contracts, and

token economies in modifying gross inappropriate behaviors of delinquents.

The Achievement Place program has been designed both to establish appropriate behavior as well as to reduce delinquent behavior. It has been a community based, family style treatment program that has incorporated a token economy and other behavior modification procedures. Previous research at Achievement Place has indicated that the behavior modification procedures are effective in alleviating many of the social, self-care, and academic behavior problems of the pre-delinquent youth (Phillips, 1968; Bailey, Wolf, and Phillips, 1970; Phillips, Phillips, Fixsen, and Wolf, 1971). The practicality of the Achievement Place program as a model rests on the assumption that it is possible to treat the important behavior problems of the boys in such family style programs without outside professional help. The research described here involved an attempt to treat articulation errors of two boys by arranging for their peers to carry out the speech therapy.

Articulation errors, the omissions, substitutions, and distortions of sounds in the speech of otherwise normal youths are said to occur at a higher rate than normal among delin-

¹This investigation was supported by Grant MH 16609 from the National Institutes of Mental Health (Center for Studies of Crime and Delinquency), and by Grant HD 03144 from the National Institute of Child Health and Human Development to the Bureau of Child Research and the Department of Human Development, University of Kansas. Preparation of this manuscript was partially supported by PHS Training Grant HD 00183 from the National Institute of Child Health and Human Development to the Kansas Center for Research in Mental Retardation and Human Development. Reprints may be obtained from Jon S. Bailey, Psychology Dept., Florida State University, Tallahassee, Fla., 32306.

²We gratefully acknowledge the assistance of David Draper, Department of Speech and Hearing, University of Kansas, for assistance in administering the articulation tests and in helping with an earlier version of this manuscript.

³Experiment II is based upon a thesis submitted by Gary D. Timbers to the Department of Human Development, University of Kansas, in partial fulfillment of the requirements for the Master of Arts degree.

quents (Cozad and Rousey, 1966). Long-term intensive counseling and speech therapy (Van Riper, 1954), and psychotherapy (Rousey and Moriarty, 1967) are the usual treatment procedures for articulation problems.

Rather than rely on expensive outside professional help, we have attempted to use the peer group as the speech therapists. The present study was undertaken to determine if pre-delinquents' peers could be used to modify their articulation errors with only minimal adult supervision.

EXPERIMENT I

Setting

Achievement Place was established by local citizens as a community based rehabilitation program for six to eight boys, 11 to 15 yr old, who were accumulating arrest records and who were, in the opinion of the juvenile court judge, near to being committed to the state Boys' Industrial School (thus the term pre-delinquent). The home was administered by a pair of teaching-parents professionally trained in behavior modification techniques. The youths, almost all from poverty backgrounds, earned points for engaging in social, self-care, and academic behaviors that were viewed as necessary for the youths' rehabilitation. The points were then spent by the boys on various privileges such as permission to watch television, go downtown, or ride their bikes, or the points could be exchanged for snacks or an allowance. Each of these privileges cost 1000 points per week when the first experiment was conducted. By the time that the second study was carried out, over a year later, inflation had almost doubled the cost of these privileges in points. That is, they then cost 2000 points per week. A more complete description of the token economy was presented by Phillips (1968).

Subject

The subject was a 13-yr-old male Caucasian who resided in Achievement Place. He came from an economically deprived background and had lived in the home 1 yr before the study began. His speech was notable for the proliferation of articulation errors that he made in normal conversation. Four of his most common errors were the substitution of /w/ for /l/ in double consonant blends (*e.g.*,

"fowr" for "flower"), substitution of /o/ for the final /r/ (*e.g.*, "motho" for "mother"), substitution of /f/ for /th/ where it was voiceless (*e.g.*, "fum" for "thumb"), and substitution of a glottal stop for /t/ and omission of the final /g/ when followed by /ing/ (*e.g.*, "kuin" for "cutting"). Three months before the study, he attended regular speech therapy sessions twice a week for two months with no apparent improvement. The therapist reported that she could not help him unless he "tried harder" to improve and practiced more at home. She dealt only with the /r/ phoneme articulation error.

The principal speech modifiers in the study were four other boys who lived in the home. They ranged in age from 12 to 15 yr and had similar social and cultural backgrounds to the subject but demonstrated no speech deficiencies.

Materials

The materials for the study consisted of 70 cards (4 by 5 in.), each of which constituted the stimulus (a pictorial representation of an object, person, or activity in most cases, or the printed word itself in some) for a verbal response. The cards represented approximately equal numbers of each of the four classes of errors and initially were shuffled to produce a random series of words. Once selected, this series was used throughout the experiment. These four classes of words were then each divided into two equal groups of words. One group of words was "treated" by the peer-trainers. The second group of words, which was never treated, served as control words for assessing possible generalization effects within each error class. All sessions were tape recorded to provide a permanent record of the subject's progress and to provide for later reliability checks.

The following list of words was used in the treated (*italicized*) and untreated classes. L: *floor, welding, telephone, flower, playing, building, clover, plumber*; stealing, island, flag, plug, climbing, blowing, sled, selling. R: *alligator, doctor, projector, hair, mother, quitter, pitcher, river, better, spider*; finger, tower, waiter, anchor, deer, beaver, pear, tear, tire, picture. TH: *thorn, thermostat, thaw, thumb, thread, thermos, threat*; thirsty, thrifty, thimble, throat, throwing, thermometer, thief. TING: *quitting, dieting, inviting, knotting,*

dating, putting, knitting, fighting, sitting, cutting; voting, heating, waiting, eating, writing, hitting, petting, resting, shooting, pointing.

EXPERIMENTAL PROCEDURES

The general experimental design for both experiments followed what is known as the multiple baseline design (Baer, Wolf, and Risley, 1968). With this design, a number of baselines of different but apparently equivalent responses are taken simultaneously. The independent variable is then applied successively to each behavior while the other continuing baselines are used as controls. This design is used when it appears either unlikely that a treated behavior will reverse when the contingencies are removed or when it is undesirable to allow the improved behavior to return to previous levels. The design was appropriate for these subjects because they each had a number of articulation errors, and it did not seem likely that once corrected the same error would appear again, nor was it viewed as desirable to reverse any improvement in their speech.

Baseline

An adult experimenter (who was involved in the testing but not in the treatment) brought the subject to a room in the home, seated him at a table in front of the microphone for the tape recorder, and showed him the 70 cards one at a time. If the subject did not emit the correct word when the card was shown (e.g., if he said "woman" instead of "mother") he was prompted to supply the correct label without the experimenter modelling the word (e.g., "No, what do you call the woman who raised you and your brothers and sisters?"). Once he could "correctly" name all the cards, the baseline sessions were begun.

During baseline, the subject was shown the 70 cards, one at a time and was asked to say each one once. He was not given any feedback on the correctness of the words, nor was he awarded any points for correct words, although he was given 100 points each day for participating in the experiment. Each session lasted from 4 to 6 min. The peers who would later correct the subject's speech were asked to attend to baseline sessions and were instructed to: "Listen to how Tom says these words, but do not say anything."

During the entire study, this 70-picture test (with no contingencies for or feedback to the subject) was given each evening to determine how many words the subject had acquired during that day's training. The data from these tests constituted the major dependent variable of the study.

Group Treatment—/l/ (negative)

Since it was assumed that a number of peers would exert more control than an individual, the first condition used the four peers together as speech modifiers. In addition, since it was desired to have them find articulation errors, any peer who judged a word to be incorrectly pronounced earned 20 points. The subject was given 100 points for participating in the training session. This was called the negative condition. At the beginning of this phase, the peers were gathered together in the game room and with the subject present, were given the following instructions:

"We are going to start a project called 'Improving Tom's Speech'. Here is a card with some words on it for each of you. (Each peer was then given a card with eight words having the consonant /l/ blend in them.) Tom is going to say a lot of words, when he says one on your card that sounds incorrect to you, and if you identify it *first*, you will have a chance to correct him on it. (In case of a tie the experimenter simply chose one of the boys to do the correction.) You will earn 20 points for each word you identify as wrong and Tom will lose 10 points. Once you hear an incorrect word and identify it, you should correct him until it sounds right to you. Then we will go on to another word. I'm not going to help you on this so you are on your own."

The experimenter showed the boys that only eight words were on the card and that they could not correct any other errors made. He then proceeded to turn over the 70 cards, stopping only when a peer identified an incorrect word and then proceeding only when the peer indicated that he judged it as correct. Points were lost by the subject based solely on the judgments of the peers. The peers were not told *how* to correct the words, nor were they told which sound in the word to correct,

even so all the peers modelled the words for the subject. At the end of the session, they tallied the number of points they earned and the subject lost.

Typically, when the subject would make an error one of the peers would shoot up his hand and yell: "Wrong!" He would then proceed to say the word correctly and have the subject repeat it after him. Often the peer modifiers would break a word into its component parts (syllables) and have the subject repeat each separately. In addition, they would target in on the particular error sound and drill on it for several repetitions, *e.g.*, "No, Tom. Say 'fla'. Now say it again. Good, now say 'flower'." It was not unusual for the peers to exaggerate the sounds and often they would correct a sound even when by the experimenter's standards it was correct. No feedback was given to the peers on any of these strategies for speech correction.

After the first three days, the peer speech modifiers conducted the sessions without any help at all from the experimenter and after the next session the experimenter was no longer present during the sessions. Audio tapes were made of all sessions.

After the session, the experimenter and occasionally an independent observer, listened to the tapes of these sessions and marked whether they agreed or disagreed with the correction of a word by the peers. A correct word response was defined as one in which the target sound of the word was clearly and distinctly present. Errors in other parts of the word were not counted. A word in which the target sound was questionable was marked as wrong. Reliability checks between the experimenter and the independent observer were made independently on randomly selected sessions.

Group Treatment—/r/ (negative)

After treatment of the consonant /l/ blend class by the peers ended, the final /r/ phoneme was treated. The group procedure was identical to the procedure for the /l/ words except that the peers were assigned 10 words to correct rather than eight.

Individual Treatment—/th/ (negative)

After correction of the /r/ words was complete, the words beginning with the voiceless /th/ were treated. Since it was possible that treatment effects might be due to the fact that

correction was carried out in front of all the peers, this procedure was changed slightly. Now, rather than having all the peers present, the correction was done by one peer at a time. Each peer was given a card with eight /th/ words on it and as before was told that he would earn 20 points for each word the subject said wrong that the peer identified and corrected. The peer was instructed to model the correct pronunciation of the word for the subject and to require the subject to say the word until he judged it correct. Since the subject then had to go through the procedure four times a day, the number of points he lost per error was reduced to 5 points. During individual treatment the subject was required to pronounce only the eight target words; he did not have to say all 70 words. Each session was recorded on tape and at the end of the session both peer and subject tallied their points.

Individual Treatment—/r/ (negative)

After the treatment of /th/ words was over, /r/ words were again treated. The procedure was the same as for /th/ except that the peers were given 10 /r/ words to correct.

Individual Treatment—/ting/ (positive)

When the peers had corrected the /r/ words, /ting/ words were next introduced. However, this time, the peers were rewarded for each word that the subject said correctly during an individual treatment session. This was called the positive condition. In this way, they would receive more points when they scored the subject's speech as correct, whereas in the negative procedure they received more points when they scored more words as incorrect. The peers were given a list of 10 /ting/ words and, whenever the subject said one of them correctly (by their judgment), the peer earned points. The peers were instructed that when the subject said a target word incorrectly, he lost points and the peer should correct it until the peer judged it to be said correctly.

Measurement of Reliability

Reliability between the experimenter and a second observer was measured for randomly selected test sessions. The second observer listened to the taped sessions out of sequence and all identifying cues were removed (*e.g.*, day or number of session). Reliability was measured by comparing the second observer's

check list with the experimenter's, word-by-word. An agreement was counted if both indicated that the subject had correctly or incorrectly articulated a given word. Per cent of reliability was calculated by dividing the total number of such agreements by agreements plus disagreements.

RESULTS

Reliability

Measurement of reliability was carried out for 12 of the daily test sessions randomly selected and listened to out of sequence by the second observer. The range of agreement was from 88% to 100% with a mean agreement of 95%.

Baseline

The baseline condition was in effect for six sessions. As shown in Figure 1, none of the /l/, /th/, or /ting/ words were correctly pronounced and only one of the /r/ words was correctly articulated during the baseline phase.

Group Treatment /l/ and /r/

Figure 1 displays the subject's behavior during the daily 70-card test sessions (these involved no feedback to the subject). Arrows indicate onset of treatment for each class and triangles indicate when each class was treated. As shown in Figure 1, on the first day of the Group Treatment, the subject correctly artic-

Articulation Errors

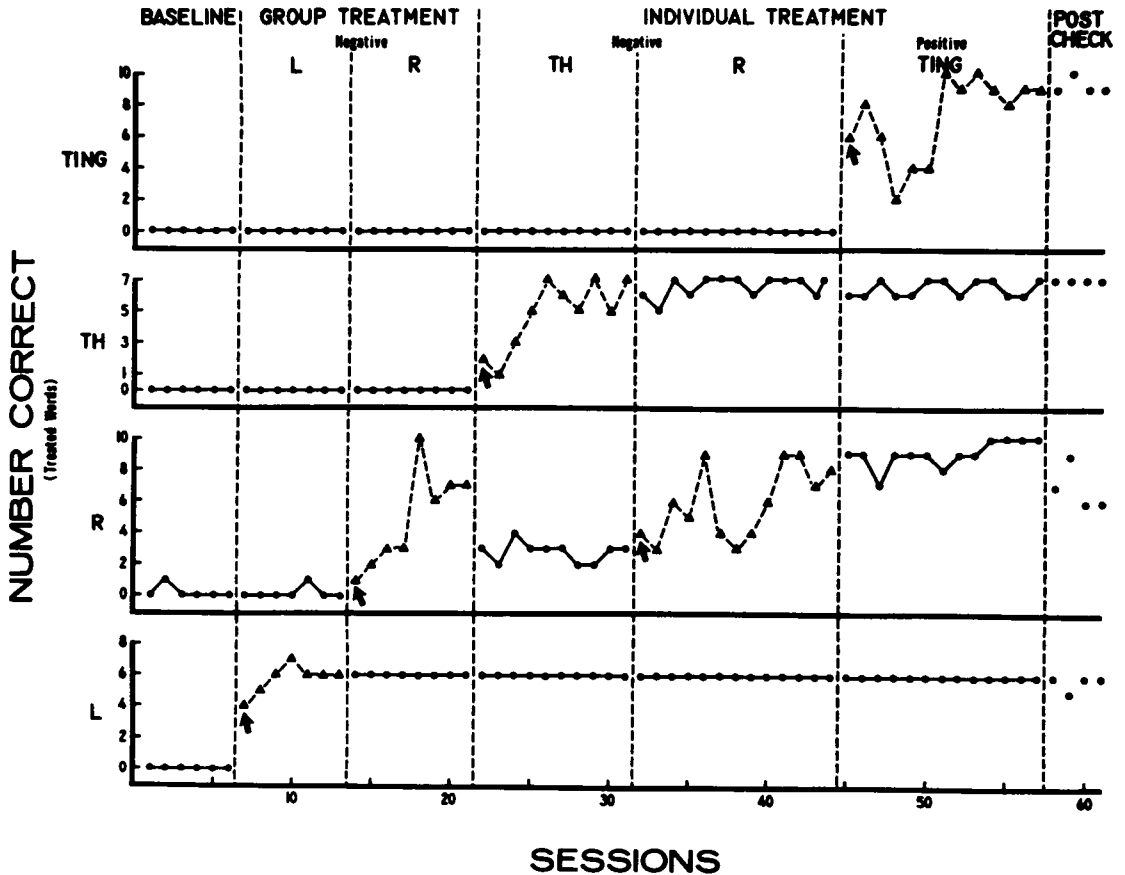


Fig. 1. The effects of peer training on four classes of articulation error in "test" sessions held each evening following treatment. Arrows mark the beginning of treatment and open triangles show when each class was treated. Group treatment was carried out by all peers together once a day. Individual treatment was carried out by each peer singly once per day. In the Negative condition, peers earned points for each word judged incorrect. In the Positive condition, they earned points for each word they judged to be correct.

ulated four of the eight target /l/ words. The subject gained one word per day for the next three days before dropping to six of eight words assigned for the final three days of this condition. During this period the subject did not make any gains in either the /th/ or /ting/ words, but did again correctly articulate one /r/ word.

After treatment of /l/s was discontinued, the same procedure was applied to the /r/s. As shown in Figure 1, relatively rapid gains were made the first five days, at which time the subject correctly articulated all 10 target words. This improvement was not maintained, however, and he stabilized at seven of the 10 words. It can be seen in Figure 1 that no changes occurred in the /th/ or /ting/ classes during this time and the improved articulations made in /l/ were maintained.

Individual Treatment /th/, /r/, and /ting/

The procedure was similar to that above except that the peers worked with the subject individually. As shown in Figure 1, once again, when the peer correction was applied, rapid improvement in articulation resulted. All seven of the target /th/ words were correctly articulated by the fifth day of individual treatment. It may be seen that no change occurred in the only remaining untreated class /ting/, while the improvement in /l/ was maintained. The /r/ words, however, showed a loss to only three correct words, dropping from seven gained initially with group treatment. The individual treatment procedure was then applied to this class of words.

When the individual treatment was applied, rapid improvement was followed by an equally rapid loss of correct articulation in the first seven sessions. Steady improvement was shown thereafter and after a total of 13 sessions the subject was correctly articulating between seven and nine of the 10 target /r/ words. At the same time, the gains in /l/ and /th/ persisted with /ting/ showing no effect.

The last class of words to be treated by the peers was the class of words containing the /t/ before /ing/, which the subject mispronounced by replacing the /t/ with a glottal stop. This time, however, the peers earned points according to how well the subject did, *i.e.*, they earned points for each word that they judged that he correctly articulated. This procedure too showed rapid gains in articulation when

first applied to the /ting/ words, with six of the 10 words being correctly articulated in the first day. This improvement was short-lived, however, and a total of 13 sessions was necessary to sustain an improvement in nine of 10 of the target /ting/ words. At the same time, the /l/ and /th/ articulations remained constant and the /r/ improved slightly again, as shown in Figure 1.

Generalization

As described earlier, each of the four classes of treated words was matched with a set of untreated or control words to be used to test for generalization of any effects observed. The data on the generalization from treated to control words for each class of error are presented in Figure 2. As in Figure 1, arrows indicate onset of correction by peers and triangles show the number of sessions that the treatment was carried on for the treated words. It may be seen that, in general, whenever the treated class was being dealt with, some generalization to control words occurred.

The improvement was considerably less than for the treated words, amounting to about 40% of the words for the /l/, /r/, and /th/ classes. Considerable improvement was shown initially in the /ting/ control words, but it rapidly fell off and very little generalization (20%) was shown toward the end of the treatment of the /ting/ class. The small improvements that did occur in the control words occurred when onset of correction for the matched treated words was begun, except for the /r/ control class, which showed some improvement before and after treatment of /r/ words.

Post-checks

One month after the above manipulations were carried out, two post-checks were given on successive days. One month after this, two more post-checks were again given on successive days. As shown in Figure 1, all classes of corrected words persisted with the exception of the /r/s.

A post-check on the generalization of effect to the control word classes was included with the post-check for treated words previously described. At this time, /r/ and /th/ words had changed little from the last session, as shown in Figure 2. Improvement can be seen in the /ting/ words. The large improvement

Articulation Errors

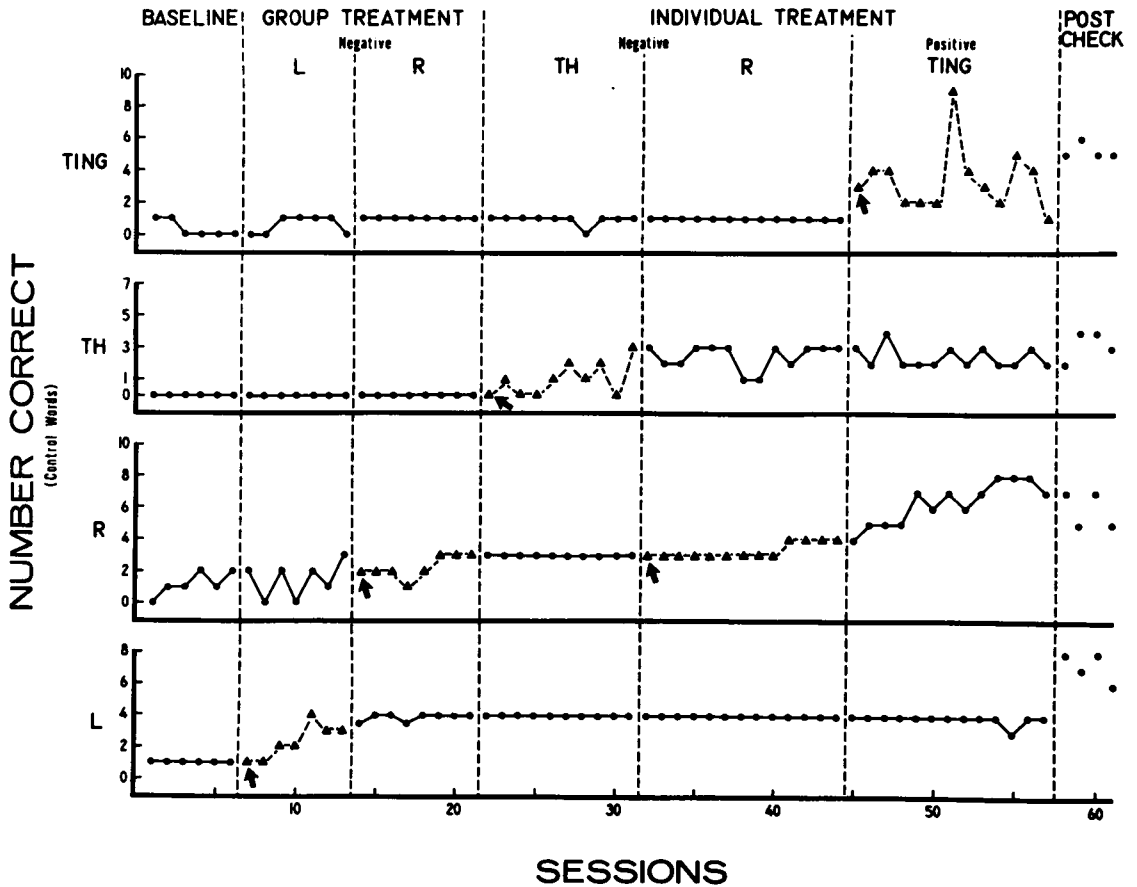


Fig. 2. Generalization of the effects of training on control words that were never treated. Arrows indicate when treatment began on the corresponding error class. Open triangles show when the corresponding class was treated.

in the /l/ class on the post-check was due to an experiment using the control /l/ words that took place between the end of the experiment and the post-check, which is described below.

Positive vs. Negative Point Conditions

The group and the individual modification condition, as well as the positive and negative point conditions, seemed to be about equally effective in motivating the peers to correct the speech errors. However, as shown in the first part of Figure 3, it appeared that the negative point condition (where the peers were given points for each error detected and corrected) caused more discrepancy (where the peer was given points for each word he judged as said correctly by the subject). That is, the peer-

trainers detected many more errors under the negative point condition than the experimenter, but about the same number under the positive point condition. To determine if this effect was reliable, a short experiment on this one variable was performed.

The experiment involved *untreated* /l/ class as the target words. Each peer worked with the subject individually. Each peer was told to correct the eight /l/ words and that he would be given 20 points for each word the subject said *correctly* (in their peer's judgment). After 11 days of training, each peer was then instructed that he would earn 20 points for each error he found and corrected (*i.e.*, the negative point condition). This condition was in effect for five days and then the positive condition

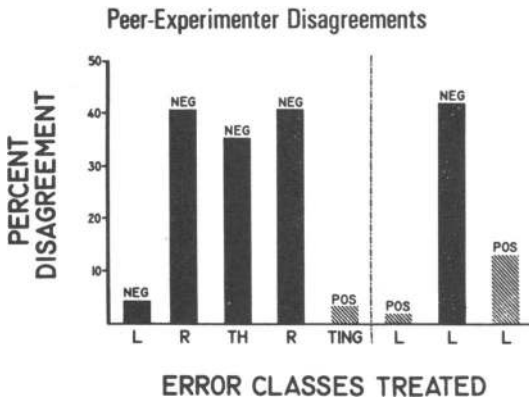


Fig. 3. Per cent of disagreement about the accuracy of the subject's articulation between the peer-trainers and the experimenter for each word class. Peers received the points for each word judged incorrect under the Negative condition and earned points for each word judged correct under the Positive condition. Peers judged more incorrect than the experimenter under all conditions.

was reinstated for six days. The experimenter monitored all training session tapes and judged whether each target word was correctly articulated or not and whether he agreed with the peer on its correctness. As before *no* feedback was provided to the peer trainers.

RESULTS

Data from the individual training sessions are presented in Figure 3 showing the discrepancy in agreements between peers and the experimenter during the negative and the positive conditions of the experiment. The peers judged more words as incorrect under *all* conditions than did the experimenter. Thus, Figure 3 was calculated by subtracting the per cent judged incorrect by the experimenter from the per cent judged incorrect by the peers. It can be seen that the positive point condition with the /ting/ class produced much closer agreement than did the previous three negative point conditions. In the next condition, when peers earned points for detecting correct /l/ responses on the part of the subject, there was a disagreement on only 1.8% of the words. When they subsequently earned points for detecting /l/ errors, the disagreement rose to 43%. In the last phase, when the positive point condition was again instituted, disagreement dropped to only 13.4%. It appeared, then, from the results shown in Figure 1, that

the two conditions (the peers earning points for detecting correct responses or incorrect responses) were about equally effective in modifying the subject's articulation errors. On the other hand, the positive condition produced much closer agreement with the observer. Thus, the positive condition may produce judgments on the part of the peers that more closely correspond to the adult world. Therefore, the positive condition is probably the more preferable of the two conditions.

EXPERIMENT II

Approximately 1 yr after Experiment I was carried out, a second youth with a severe articulation disorder entered the home. It was decided to replicate Experiment I with this youth. In addition to the replication of Experiment I, Experiment II was designed to provide several methods of testing generalization, and further to analyze the effects of different reinforcement conditions upon the peer-speech-trainers.

Subject

A 12-yr-old male Caucasian, from an economically deprived background, had been classified as dependent neglected by the juvenile court. Five other boys who lived in the home, all of whom had similar backgrounds but had more normal speech, served as the speech trainers.

The subject's articulation errors fell into four classes with each class having a letter sound that, because of an omission, insertion, or distortion, was incorrectly pronounced. An initial /z/ phoneme was pronounced as /j/ (e.g., "jipper" for "zipper"). Errors in /l/ words were of two varieties. The subject omitted the /l/ phoneme in words containing double consonant blends (e.g., "probem" for "problem"). A second, less prevalent, error was the omission of the medial /l/ (e.g., "faiure" for "failure"). In words ending in *ing* the subject omitted the final /g/ (e.g., "lookin" for "looking"). In /r/ words the medial /r/ sound was omitted (e.g., "ion" for "iron"), or the /o/ sound was substituted for the terminal /r/ (e.g., "featho" for "feather").

Eighty four words containing the above described articulation errors were collected. There were 9 /z/, 23 /l/, 25 /ing/, and 27 /r/

words. Words were chosen that were consistently mispronounced by the subject before the study began. The response definition for determining the correct pronunciation of a word was that the target sound could be distinctly heard and was correctly pronounced. In addition, the word was judged as correct or incorrect only upon the pronunciation of the specific sound occurring in the word and not upon the pronunciation of the whole word. When more than one of the above mentioned errors occurred in a word, only one was designated as the target error sound. For example, the word "cradle" could have both the /r/ and the /l/ sound that could be articulated incorrectly, but only one sound was chosen to be a target sound.

Each class of error words was divided into two groups. One group of words were the "treated" words. These were worked with in the training sessions. The second group, the "untreated" words, were never worked with and served as a test for generalization of the effects of training within each class of words. The words were randomly divided into the treated and untreated groups.

The following is a list of the words that were used in the treated (*italicized*) and untreated classes. Z: *zebra, zillions, zero, zoo, zone*; zoom, zinc, zig, zipper. L: *flame, calm, explore, Carlton, sold, world, failure, volume, welding, told, paddle, bold*; problem, coil, field, cold, cradle, hold, blame, old, silver, oil, fold. ING: *sweeping, singing, cheating, washing, loving, swimming, smoking, walking, quitting, blowing, snowing, forgetting, sleeping*; checking, watching, cooking, listening, fetching, throwing, baking, dieting, disobeying, fishing, looking, clothing. R: *together, ever, scouring, behavior, write, iron, leather, cereal, spider, library, mayor, hunger, feather, bother*; tire, warm, cover, brother, arm, finger, worm, cheer, teacher, beaver, radar, deer, ruler.

Materials

The words were presented to the subject on a card in the form of a picture. Pictures rather than printed words were used as stimuli to prevent any cues for pronunciation that the printed words might provide. The pictures were drawn by the subject on 4 by 6 in. cards. The subject was instructed how to tape record his own test and treatment sessions. The test sessions usually lasted about 10 min each.

Baseline

A baseline measure of the speech behavior was obtained during test sessions for all of the 84 words. The subject was told to pronounce the words that the pictures on the cards represented. The experimenter was present during the 10 baseline test sessions but not thereafter. Three hundred points were earned by the subject for participating in the baseline test sessions and no feedback was given as to the correctness of the articulations.

Also during baseline, two standard articulation tests were administered by a trained speech pathologist. The Templin-Darley Tests of Articulation contained 141 items (Templin and Darley, 1969). Included in the test were items testing 34 /r/, 19 /l/, 4 /z/, and 2 /ing/ sounds. The McDonald Deep Test of Articulation enabled testing of the /r/, /l/, and /z/ phonemes with 49 test items for each sound (McDonald, 1964). The tests were administered imitatively because of the subject's poor reading ability. The same two tests were administered a second time at the conclusion of all peer speech training.

Treatment Procedures

The general training procedures for each class of treated words were as follows. In individual sessions, the subject would pronounce one word at a time for the peer. Fifty points were earned by the subject if the peer judged the word to be correct. He lost 50 points for words judged by the peer as incorrect. In addition, after an incorrect response, the peer was instructed to model the correct pronunciation and the subject was told to try again. This pattern was repeated until the subject was judged by the peer to be correct or to have made at least some progress toward a correct response. Points were either earned or lost for the first pronunciation of each word and not for any subsequent responses during the training. The entire training process was carried out for each of the treated words. Each of four or five boys (depending upon the number of boys in the home at the time) gave one training session a day. The subject earned 50 points for participating in each of the sessions and each peer earned 100 points. As in the first study, the peers were given no instructions about speech training or supervision or feedback about their training procedures. The

peers were not told what the target errors were or how the words should be correctly pronounced, except in one condition, which will be described later. The experimenter was not present during the training sessions.

Treatment /z/ and /r/ (positive)

There were two reinforcement conditions for the peer-speech-trainers. In the positive condition, 50 points were earned by the peer each time he judged the subject's first pronunciation of a word as correct. The positive condition was in effect for the training of the /z/ and /r/ words, and for the second training phase of the /l/ words.

Treatment /l/ and /ing/ (negative)

In the negative condition, the trainers earned 50 points each time they judged the first pronunciation of a word by the subject as incorrect. The negative condition was used for the training of the /ing/ words and for the first, third, and fourth phases of the /l/ words. The fourth phase of the /l/ training involved only four /l/ words that had not been pronounced correctly in any of the preceding test sessions. While no pronunciation training had been given before, the experimenter instructed the peer on the proper pronunciations of these four words before the last /l/ training phase began. These instructions consisted of the experimenter modelling the correct pronunciations of the four words and then teaching the peers to correctly pronounce the four words. These instructions lasted for 5 min and were never repeated.

Daily Test

At the completion of all training sessions for one day a test session on all 84 words was carried out. The procedure was the same as for the baseline sessions, except only 50 points were earned by the subject for his participation. After the baseline sessions, the experimenter was not present for the daily tests.

Sentence Generalization Test

To approximate the use of the words in spontaneous conversation, the subject was periodically requested to use all the words in sentences. This was called a sentence generalization test. Rather than use the stimulus cards, the experimenter would either give a definition of the word or point to an object as

a cue for the word. The subject would then make up a sentence using the word. Following this procedure, the subject would use all the words in sentences. The subject received 300 points for his participation in the sentence generalization tests but was given no feedback as to the correctness of his pronunciations of the words.

Measurement of Reliability

Measurement of agreement between independent observers was carried out for the experimenter's scoring of the words in the test sessions, sentence generalization tests, and training sessions. A second scorer, unaware of the words being trained or of the experimental conditions in effect, carried out the reliability checks. Reliability was calculated by comparing the scorings item by item and dividing agreements by agreements plus disagreements.

Test sessions were tape recorded by the subject on a Wollensak Stereo Recorder Model #5710. All scoring by the experimenter, as well as reliability checks, were taken from tapes.

RESULTS

Reliability

Inter-observer agreement averaged 91% for 17 reliability checks carried out on the pronunciation of the 84 words in the test sessions. A further analysis showed that agreement calculated for the treated words was the same as for the untreated words (91%). Three reliability checks were carried out on the sentence generalization tests. Inter-observer agreement averaged 91% for all words. For treated and untreated words used in sentences, agreement averaged 89% and 92% respectively.

Treated Words

Ten baseline test sessions were carried out, and as shown in Figure 4, there were no treated words pronounced correctly in the /z/ and /l/ classes. In any one test session, no more than three /ing/ and two /r/ words were pronounced correctly during the baseline test sessions.

Treatment /z/ (positive)

Training of the /z/ class was carried out with the positive condition in effect. As shown in the second phase of Figure 4, all /z/ words were articulated correctly in the test session

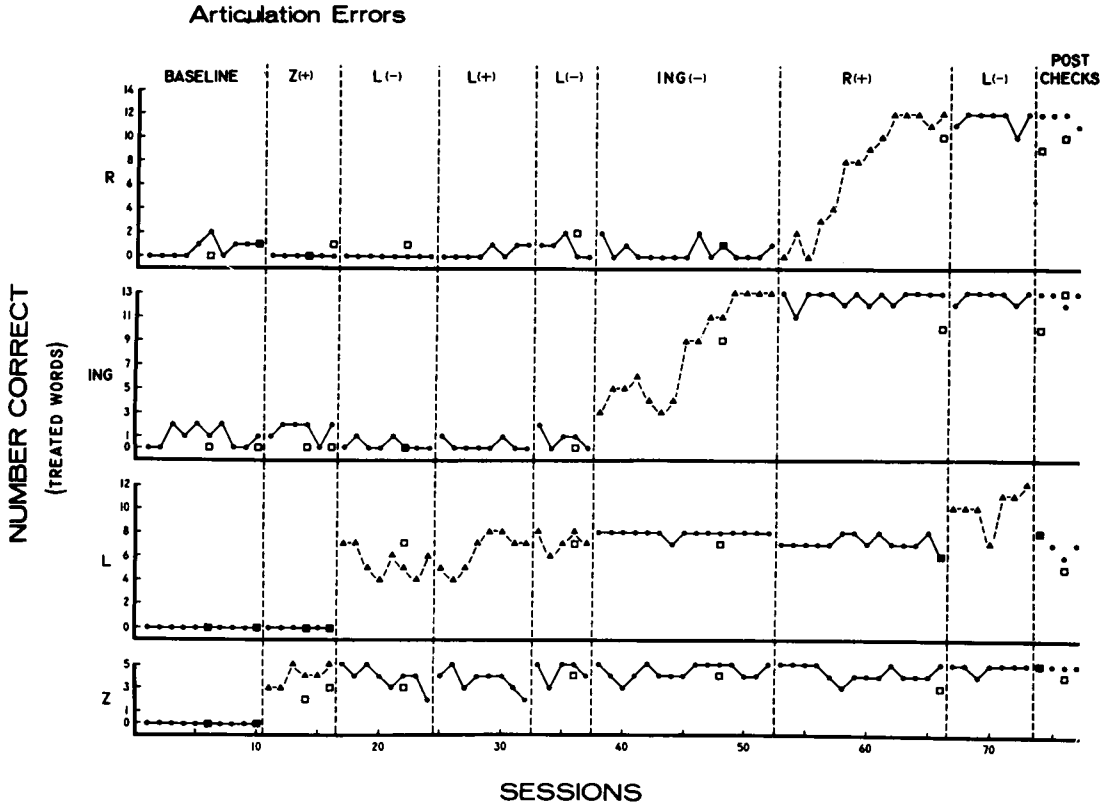


Fig. 4. Number of words correctly pronounced under each training condition. The filled circles represent the number of treated words that were pronounced correctly in the daily test sessions. The triangles indicate the class of words being trained. Squares indicate sentence generalization tests.

after three days of training. In the test session following the sixth day of training, all /z/ words were again pronounced correctly and training of the /z/ words ceased. During training of the /z/ class, no improvement occurred in the /ing/ class and no /l/ or /r/ words were articulated correctly:

Treatment /l/ (positive and negative)

Between Sessions 17 and 37, the /l/ class of words was trained successively with the negative, positive, and negative conditions in effect for the peers. As shown in the third phase of Figure 4, there were seven correct pronunciations of the /l/ words in the first test session after training began. After the second session, a downward trend can be seen, but, it reversed three sessions later. The negative training condition was in effect for eight days followed by the positive condition for eight days. As the fourth phase of Figure 4 shows, there was an increase to eight correct pronunciations during the positive condition for /l/

words. As shown in the fifth phase of Figure 4, the last five training sessions carried out with the negative condition resulted in no increase in the number of correct pronunciations in the test sessions. During training of the /l/ words, the /ing/ and /r/ words continued to be misarticulated, and, the number of correct articulations did not rise above baseline level. At the same time, the /z/ words continued to be correctly articulated at the rate reached during training.

Treatment /ing/ (negative)

The /ing/ words were trained for 12 days with the negative condition in effect. As shown in the sixth phase of Figure 4, in the test sessions following the last four training sessions, all 13 of the /ing/ words were correctly pronounced. While training of the /ing/ words was carried out, the /z/ and /l/ words continued to be articulated correctly at the level they reached during their respective training phases.

Treatment /r/ (positive)

The /r/ words were trained for 14 days with the positive condition in effect for the peer-speech-trainers. As seen in the seventh phase of Figure 4, after the tenth day of training, 12 of 14 of the /r/ words were articulated correctly in the test session. The /z/, /l/, and /ing/ words maintained the level of accuracy they had previously reached.

Treatment /l/ (negative)

The last phase of training involved four treated /l/ words that the subject had previously not pronounced correctly in the test sessions. At the beginning of this phase the peers received special instructions on the correct pronunciation of the four words. These instructions lasted for about 5 min; thereafter, the peers began working with the subject on the four words. The peer training lasted for seven days and was carried out with the negative condition in effect. As shown in the eighth phase of Figure 4, in the test session following the seventh training session, all four

of the /l/ words were pronounced correctly along with the remainder of the treated /l/ words that had been previously trained.

Sentence Generalization Tests

As shown by the squares in Figure 4, the number of treated words correctly pronounced in the sentence generalization tests corresponded closely to the results of the daily test sessions. In the sentence generalization tests before the training of each class of words, the number of words correctly pronounced was near zero. After training was carried out on each class of words, the number of correct pronunciations increased to an average of 65%, but the accuracy level was generally lower for the words when they were used in sentences than when they were pronounced individually in the test sessions.

Generalization

As shown in Figure 5, during the 10 baseline test sessions the subject did not pronounce any of the untreated /z/ words correctly. There was one /l/, three /ing/, and two /r/

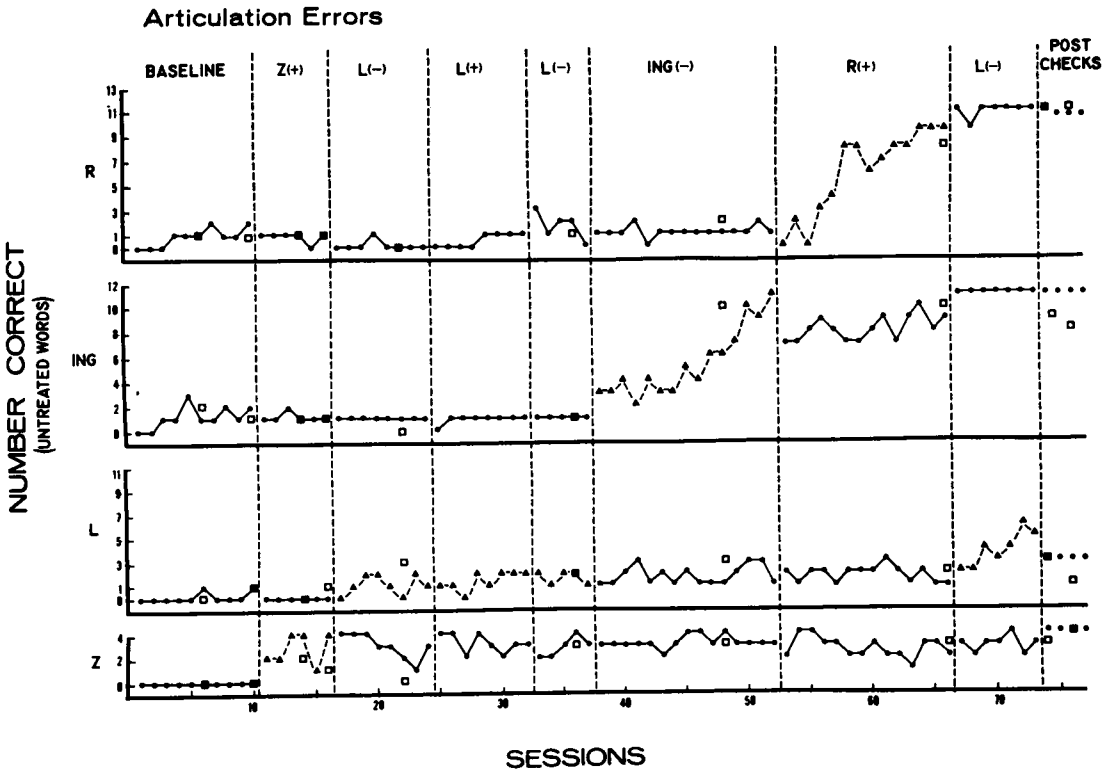


Fig. 5. Effects upon the pronunciation of untreated words occurring when training was carried out on the corresponding class of treated words. The filled circles represent untreated words and triangles indicate when the corresponding class was trained. Squares indicate sentence generalization tests.

words pronounced correctly during the baseline test sessions.

As training of the treated /z/ words was carried out, all four of the untreated /z/ words were pronounced correctly after the third training session. During training of the /z/ words, the untreated /ing/ and /r/ words remained within the baseline range and there were no correct pronunciations of the untreated /l/ words.

There were three phases of training for the treated /l/ words. As the three training phases were carried out, there were two untreated /l/ words correctly pronounced in the test sessions. At the same time, the untreated /z/ words maintained the level they reached during training of the treated /z/ words. The untreated /ing/ and /r/ words remained within the baseline level except in test session 33, when there were three correct /r/ pronunciations.

As training of the treated /ing/ words was carried out, the number of untreated /ing/ words correctly pronounced increased to 11 of 12 in the test session after the last day of training. At the same time, the untreated /z/ words maintained the level that was previously reached and three of the untreated /l/ words were correctly pronounced during several of the testing sessions. The untreated /r/ words remained within the baseline level.

There were nine of 13 untreated /r/ words pronounced correctly in the test session after 12 sessions of training the treated /r/ words. At the same time the untreated /ing/, /l/, and /z/ words maintained the level of accuracy they had previously reached.

After the special training condition of the four treated /l/ words, six untreated /l/ words were pronounced correctly in the test session following the sixth day of training. While the training of the /l/ words was carried out, the subject maintained the untreated /r/, /ing/, and /z/ words at the level of accuracy he had previously reached. The average improvement in the untreated words across all these conditions was 78%.

Sentence Generalization Test

As shown by the squares in Figure 5, the number of words correctly pronounced in the sentence generalization test corresponds closely with the number of correct pronunciations occurring in the test sessions. The improve-

ment averaged 55%. The correspondence was apparent both before and after the training of the treated words in each class.

Post Checks

A total of four post checks was carried out. These included four test sessions and two sentence generalization tests. The last post check was taken a little over two months after the last training session had occurred. As Figures 4 and 5 indicate, the level of accuracy reached during training was maintained for the treated and untreated /z/, /ing/, and /r/ word classes. The level of accuracy for the treated /l/ words decreased to the level that it had been before the last special training condition for the four /l/ words. The increase of the untreated /l/ words obtained during the special /l/ training condition was not maintained either. The post checks of the sentence generalization tests indicated that the accuracy level had been maintained for both the treated and untreated words.

Standard Articulation Tests

Figure 6 shows the results of the McDonald Deep Tests of Articulation and the Templin-Darley Tests of Articulation. Each test was given before the speech training and at the completion of all training. Significant improvement was seen for most categories. As seen in the upper graph of Figure 6, on the McDonald test, improvement was from 41% to 80% for the /r/ category and 47% to 97% for the /z/ category. Significant improvement was restricted in the /l/ category because of the subject's high performance on the pre-test. He did improve his performance, however, from 92% to 100%. As the lower graph of Figure 6 shows, the improvement on the Templin-Darley test was 38% to 68% for the /r/ category and 50% to 100% for the /ing/ and /z/ categories. On this test too, the subject did well on the pre-test in the /l/ category. His improvement in this category was from 89% to 100%. Reliability checks were carried out on the second administration of each standard articulation test. A trained speech pathologist was the second observer. Per cent agreement averaged 89% on the McDonald Deep Tests of Articulation and 89% on the Templin-Darley Test of Articulation.

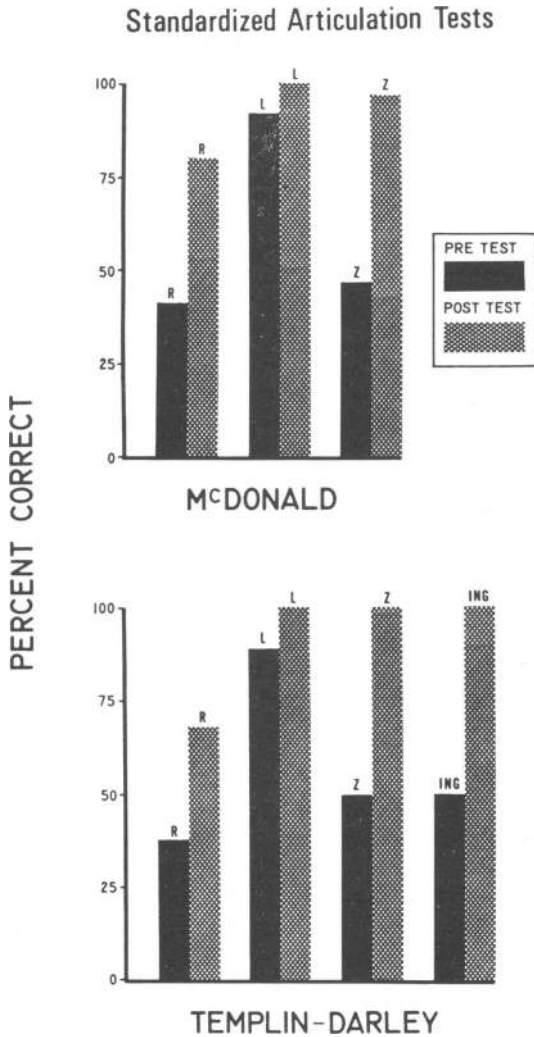


Fig. 6. Results of pre and post tests of the McDonald Deep Tests of Articulation and the Templin-Darley Test of Articulation.

Positive and Negative Training Conditions

Figure 7 shows the results of the training sessions. The triangles represent how the peers judged each of the treated words that were trained in each session. Each point represents an average of the four or five sessions that were carried out each day. (The number of sessions depended upon the number of boys in the home.) The solid line represents experimenter's score of each of the training sessions. The experimenter scored the words from the tapes of the sessions. The peers never received any feedback as to the experimenter's scores. The numbers along the horizontal axis correspond

to the number of the test session that was carried out at the completion of all training for a day.

The /z/ class of words was treated first and the positive condition was in effect for training these words. That is, the peers earned points for every "correct" response they judged the subject to emit. As seen in the first phase of Figure 7, during the first two training sessions the peers judged more responses as correct than did the experimenter. For three of the last four sessions the peers judged more responses as incorrect than did the experimenter.

The first phase of training for the /l/ words was done with the negative condition in effect for the peers. Therefore, the peers earned points for every "incorrect" response they judged the subject to emit. As shown in the second phase of Figure 7, for the first training session, and for the last three training sessions, the peers scored more of the subject's responses as incorrect than did the experimenter. The opposite was true for the middle four training sessions.

For the next eight training sessions, the /l/ words were trained with the positive condition in effect. As indicated in the third phase of Figure 7, in all training sessions, the peer judged more responses as correct than did the experimenter. For these sessions the peers judged between 82% and 98% of all the subject's pronunciations as correct. For the sessions, the experimenter judged between 38% and 78% of the subject's pronunciations as correct.

Finally, for the next five training sessions, the /l/ words were again trained with the negative condition in effect. As shown in the fourth phase of Figure 7, the peers judged more responses as correct than did the experimenter. However, in contrast to the previous positive condition, the peers found fewer correct responses. The number of pronunciations judged as correct by the peers ranged between 64% and 86%. For the same sessions the experimenter judged between 61% and 72% of the pronunciations as correct.

Training of the /ing/ words was carried out with the negative condition in effect. In the fifth phase of Figure 7, except for the first two training sessions, the peers continually judged more of the subject's responses incorrect than did the experimenter.

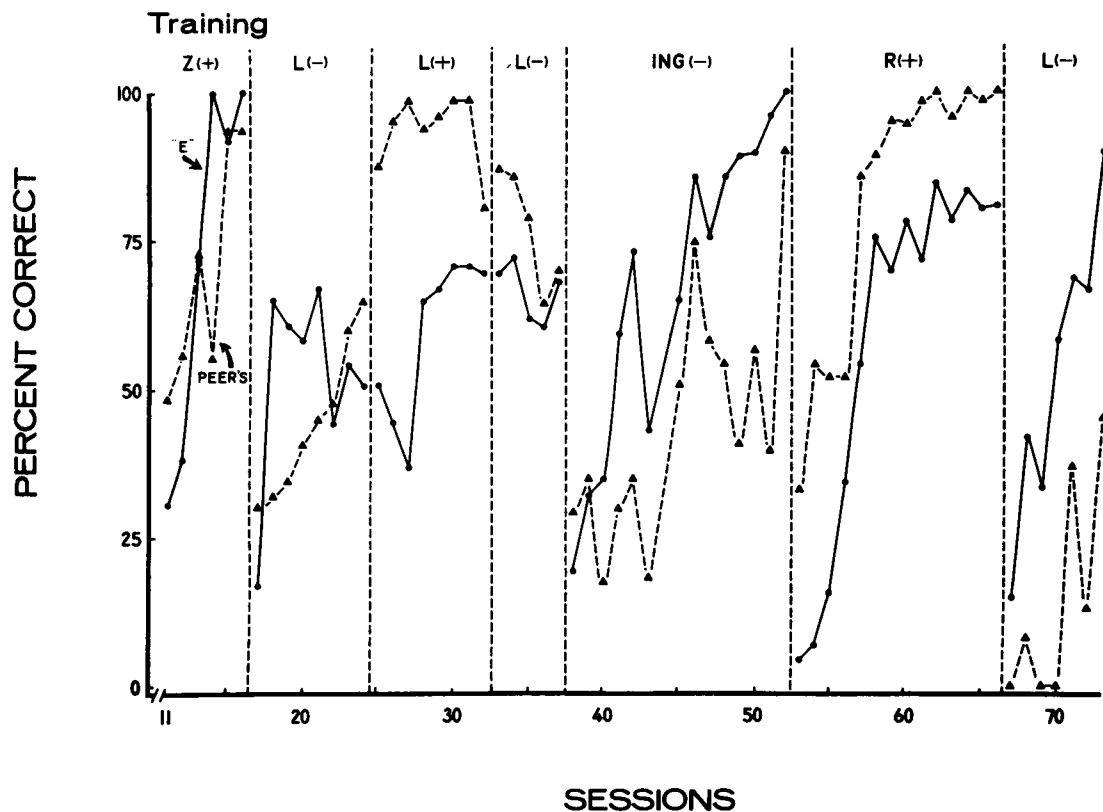


Fig. 7. Results of training sessions. The triangles represent a daily average of how the peer-speech-trainers scored each of the words being trained. Filled circles represent how the experimenter scored the same words.

The positive condition was in effect for the training of the /r/ words. As the sixth phase of Figure 7 indicates, for all sessions the peer-speech-trainers judged more of the subject's pronunciations as correct than did the experimenter.

As shown by the seventh phase of Figure 7, when the four /l/ words were trained with the negative condition in effect, the peers in all sessions judged more pronunciations as incorrect than did the experimenter.

In summary, there tended to be a bias on the part of the peers. Their judgments were biased in the direction in which points could be earned. Typically, in the positive condition, when the peers earned points for judging a response as correct, they would judge more responses as correct than did the experimenter. Conversely, in the negative condition, when the peers earned points for judging a response as incorrect, the peers judged more responses as incorrect, than did the experimenter.

Reliability checks by a second observer were carried out with tapes of eight of the train-

ing sessions. Reliability was calculated by the method discussed previously. Inter-observer agreement averaged 82%. For these same eight sessions, reliability was calculated between the experimenter and the trainers and between the independent observer and the trainers. Agreement averaged 58% and 59% respectively between the experimenter and the peers and the independent observer and the peers.

GENERAL DISCUSSION

The results of Experiment I indicated that a delinquent's peers could successfully be used as speech modifiers for articulation errors. In this study the peers were able to improve the subjects articulation with almost 90% of the words they were given to train. In addition, an improvement of approximately 40% was seen in matched control words that were never treated, indicating that some generalization of training was occurring. Further, these improvements were possible by giving the peers no speech training and almost no adult

supervision. That is, after brief instructions the peers carried out the treatment without the experimenter present. They delivered their own points and got no feedback even after the session was over. Group and Individual treatment methods proved to be about equally effective, but the positive condition where the peers earned points for each response by the subject that they judged as "correct" produced closer agreement with the experimenter than did the negative condition. That is, in the negative condition the peers would frequently judge as incorrect a word the experimenter judged as correct. They would then proceed to "overtrain" on the word. Improvements in articulation maintained after two months of training indicated the durability of the training effects.

Experiment II replicated these findings with a second subject who had similar articulation difficulties and with new peer-group members as trainers. Improvement was seen in 86% of the words the peers were given to train and an additional improvement in 78% of the control words was also evident. In addition, it was shown that the improvements in articulation generalized to the correct use of the words in sentences (a test condition that approximates normal speech). Peers required training on only one class of words to accomplish these results. Further, the subject improved significantly on commonly accepted standard speech tests, indicating that the training was of measurable therapeutic value. However, in contrast to Experiment I, both the positive and negative conditions produced equally close agreement with the experimenter. Even so, the peers did average a higher proportion of judgments of "correct" when they received points for finding "correct" responses as opposed to when they received points for detecting "errors" by the subject. Finally, each of these conditions seemed to be about equally effective in training the correct articulations.

These studies showed the usefulness of the multiple baseline design for demonstrating experimental control in situations where reversal designs would be inappropriate. In the present studies, once improvement in a specific articulation class was made by the subject as a result of peer training, the training for that class ceased. With one exception (/r/ in Experiment I), the gains did not diminish when training stopped. Thus, the necessity of an individ-

ual subject design that takes irreversibility into account is evident. There is always the risk in using the multiple baseline design that the multiple responses may not be functionally equivalent and thus the effects may not be replicated with each response. However, this was not the case in these studies.

Recent research in speech therapy has pointed up the potentially significant role that response consequences may play in the modification of disfluent speech. Point loss (Siegel, Lenke, and Broen, 1969), the word "wrong" (Martin and Siegel, 1966a; Quist and Martin, 1967), and mild electric shock (Martin and Siegel, 1966b) when made contingent on stuttering have been shown to reduce this form of disfluency greatly. In the present studies, not only were misarticulations followed by point loss but also by a peer modelling the correct articulation. Thus, in the present results it is not possible to separate the effects of these two variables or of other variables such as peer approval or feedback. However, modelling of correct articulations is standard practice in speech therapy (Van Riper, 1954), and this alone did not seem to prove helpful with the present subjects when they attended regular speech therapy sessions before the experiment began. These results, then, suggest that the addition of some form of response cost for articulation errors plus modelling may well aid in the modification of this form of speech disfluency just as response cost has been demonstrated to be extremely powerful in modifying stuttering.

The results of these two studies also have other implications. First, they suggest that peers may be quite effective in modifying the behavior of other youths with behavioral deficits. These results support the Achievement Place Model (Phillips, 1968; Bailey, Wolf, Phillips, 1970) of community based treatment for juvenile offenders. The Achievement Place model is structured such that the only professional personnel required for its maintenance are the teaching parents. It is their explicit purpose to educate the youths of the home in a variety of social, self-care, academic, and pre-vocational skill areas. With the emphasis, therefore, on the home as a self-contained unit, it becomes feasible to look at the youths as an untapped resource. Not only can the youths aid in the physical maintenance of the home (Phillips, 1968), but they can also serve as

trainers of their peers. This conclusion is supported by the present research in which the youths served as effective speech modifiers. Other areas need to be investigated in which the youths can effectively train their peers. These areas may include the teaching of social interaction skills such as appropriate greetings, conversation, and manners and academic skills such as the tutoring of reading, math, or spelling.

The notion of peers-as-trainers may also be extended to other educational settings. Schools may discover that older or high-achieving children can easily teach younger or low-achieving children certain routine academic subjects, thus relieving the teachers to work intensively with special learning problems requiring a professional teachers' training.

REFERENCES

- Baer, D. M., Wolf, M. M., and Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, 1, 91-97.
- Bailey, J. S., Wolf, M. M., and Phillips, E. L. Home-based reinforcement and the modification of pre-delinquents' classroom behavior. *Journal of Applied Behavior Analysis*, 1970, 3, 223-233.
- Burchard, J. and Tyler, V. The modification of delinquent behavior through operant conditioning. *Behaviour Research and Therapy*, 1965, 2, 245-250.
- Cozad, R. and Rousey, C. Hearing and speech disorders among delinquent children. *Corrective Psychiatry and Journal of Social Therapy*, May 1966, 12, 250-257.
- Martin, R. R. and Siegel, G. M. The effects of simultaneously punishing stuttering and rewarding fluency. *Journal of Speech and Hearing Research*, 1966, 9, 466-475. (a)
- Martin, R. R. and Siegel, G. M. The effects of response contingent shock on stuttering. *Journal of Speech and Hearing Research*, 1966, 9, 340-352. (b)
- McDonald, E. T. *A deep test of articulation*. New York: Stanwix House, 1964.
- Phillips, E. L. Achievement Place: token reinforcement procedures in a home-style rehabilitation setting for pre-delinquent boys. *Journal of Applied Behavior Analysis*, 1968, 3, 213-223.
- Phillips, E. L., Bailey, J., and Wolf, M. W. *Achievement Place: a token economy in a home-style rehabilitation program for juvenile offenders*. Paper presented at the meeting of the American Psychological Association Convention, Washington, D.C., September, 1969.
- Phillips, E. L., Phillips, E. A., Fixsen, D. L., and Wolf, M. M. Achievement Place: modification of the behaviors of pre-delinquent boys within a token economy. *Journal of Applied Behavior Analysis*, 1971, 4, 45-59.
- Phillips, E. L., Wolf, M. M., Bailey, J. S., and Fixsen, D. *The Achievement Place model: Community-based, family-style, behavior modification programs for pre-delinquents*. Paper presented to the Delinquency Prevention Strategy Conference, Santa Barbara, California, February, 1970.
- Quist, R. W. and Martin, R. R. The effect of response contingent verbal punishment on stuttering. *Journal of Speech and Hearing Research*, 1967, 10, 795-800.
- Rousey, C. and Moriarty, A. E. *Diagnostic implications of speech sounds*. Springfield, Ill.: Charles C. Thomas, 1965.
- Schwitzgebel, R. *Street corner research*. Cambridge: Harvard University Press, 1964.
- Templin, M. C. and Darley, F. L. *The Templin-Darley tests of articulation*, Bureau of Educational Research and Service, State University of Iowa, 1960.
- Thorne, G. L., Tharp, R. G., and Wetzell, R. J. Behavior modification techniques: New tools for probation officers. *Federal Probation*, 1967, 31, 21-27.
- Tyler, V. O. and Brown, G. D. The use of swift, brief isolation as a group control device for institutionalized delinquents. *Behaviour Research and Therapy*, 1967, 5, 1-9.
- Van Riper, C. *Speech correction: principles and methods*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1954.

Received 4 August 1970
(Revised 23 August 1971.)