ABSTRACT

AN EXPLORATORY STUDY OF THE EFFECT
OF VARYING MODES OF POSITIVE REINFORCEMENT
ON STUDENT ANIMATION IN THE BEGINNING SPEECH CLASS

By

Dan Pyle Millar

The purpose of the research was to determine if positive reinforcement presented by varying modes can result in the improvement of the beginning speaker's speaking skill. Particular attention was directed toward light on-set reinforcement as applied to a physical aspect of the nonverbal code --- gesture. Nine intact groups, drill sections of the beginning speech course, were compared via nonparametric analysis of variance routine. Improvement in gestural skill was measured by the difference between evaluations of judging panels during the pre-test subtracted from those during the post-test.

Two dependent measures of gestural skill were created. The Gestural Evaluation Score was the total of a three-judge panel received by an Ss on nine elements of gesturing. The Physical Performance Score was the total score received by an Ss on five Dynamism scales. Each

dependent measure was assumed to be unidimensional yet related (supported by correlational data). The judging panels evaluated the speaking of the Ss before and after exposure to three modes of positive reinforcement. In the Minimum Written Reinforcement condition, Ss received an instructor's written critique of a speaking performance and an experimental evaluator's written scale evaluation of a single element of the gestural performance. In the Maximum Written Reinforcement condition, the Ss received the instructor's written critique and an experimental evaluator's written scale evaluation of nine elements of gestural performance. In the Light On-Set Reinforcement condition, the Ss received an instructor's written critique and a flash of light for each and every gesture determined acceptable to the experimental evaluator.

Reinforcement theory would lead to the prediction that the Ss receiving the immediate positive reinforcement, found only in the Light On-Set condition, would show greater improvement than those receiving either of the written reinforcements. The greater positive reinforcement, assumed extant in the Maximum Written condition, also would be predicted to generate greater change than the Minimum Written condition. The chief hypothesis of the research, then, was that the three positive reinforcement conditions ought to generate different improvement effects. If improvement was significant, then the Maximum Written should show greater

change than the Minimum Written, and the Light On-Set should show greater change than either written condition.

No statistical significance was found in the data. Trends occur in the predicted direction for the Maximum and the Light On-Set Ss. The Ss in the Minimum Written condition show a strong trend in the negative direction. While no significance was found, the trends suggest the theoretic hypothesis to be sound and that increased control within an experimental framework would encourage statistical significance in a replication.

Accepted by the faculty of the Department of Speech and Theatre, College of Communication Arts, Michigan State University, in partial fulfillment of the requirements for the Doctor of Philosophy degree.

William B. Lashbrook
Director of Thesis

Guidance Committee: William B. Lashbrook
Chairman

Gerald R. Miller

Walter F. Johnson

Leroy L. Homer

Kenneth G. Hance

AN EXPLORATORY STUDY OF THE EFFECT OF VARYING MODES OF POSITIVE REINFORCEMENT ON STUDENT ANIMATION IN THE BEGINNING SPEECH CLASS

By

Dan Pyle Millar

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Speech and Theatre

1969

ACKNOWLEDGMENTS

The writer expresses thanks to William B. Lashbrook, the thesis advisor, for his insight and direction of this project, and the members of the committee, Kenneth G. Hance, Walter F. Johnson, Gerald R. Miller and Gordon L. Thomas, for their individual suggestions and encouragement. A special note of gratitude is extended to Dennis Dunne, Thomas Droessler and Herschel Mack, who gave up much of their own study time, as well as Donna Emlich, Greg Gardner and Carol Shaw, who saw that the research was carried out. Also, the writer acknowledges his wife who performed the manual and emotional tasks which brought this project to completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

RATIONALE AND HYPOTHESES

Introduction

When the college student enters the beginning speech classroom, he has been speaking to other people for at least 18 years. The problem facing the college speech instructor is to encourage the student to organize, develop and correct habits he already possesses. Thus, the teaching of public speaking seeks two objectives: (1) the understanding and remembering of a set of theoretic principles, and (2) the refinement of individual skills. An instructor of a beginning speech course needs to discover techniques from within and without the field of Speech which will maximize the achievement of these two objectives. The focus of this research is upon the second of the objectives --- the refinement of personal skills. The particular interest is the determination of the feasibility of varying modes of positive reinforcement as means of increasing the impact of instruction in one of the fundamental areas of student performance --- animation.

Rationale

A major factor in the refinement of the performance skills of a student in the beginning speech course is the instructor's critique of that performance. The instructor

must identify weaknesses, offer constructive means for eliminating those weaknesses, and reward the removal of error. This alternating process of approval and disapproval is designed to improve the skills of the beginning speech student. In other words, the consequence of the critique of performance is the development of desirable speaking behavior. Robinson and Kerikas (1963) suggest the importance of the teacher's role in skill development when they write that criticism underlies all learning in the speech classroom. Improvement in the individual skills, particularly delivery skills, according to Densmore, rests "first, last, and always upon the teacher in charge (1946, p. 67)." Donald Smith (1961) calls the criticism function of the instructor the "central reason for being" and indicates this function has not changed with speech instruction over the centuries. With the critique having such importance, whatever may improve the ability of the instructor to communicate his evaluations and recommendations of the performance to the student should be of value in the teaching of public speaking.

Given the importance of the instructor's communications to his students, there is still need for research in this area of speech education. Donald Dedmon notes that "much research is still needed on the subject of criticism of student speeches. The possible relationship between techniques of criticism and modification of speech behavior needs to be explored (1967, p. 284)." One way to carry on such research would be to take the reinforcement

findings of the laboratory and place them within the classroom, a recommendation advanced by Resnick (1963). Concerning laboratory findings, Becker (1963) suggests that the field of Speech needs to research what events may be reinforcing to the student and how these events may be presented to the student during class periods when the student is practicing speech-making. To research the instructor's critique, his chief technique for feeding back to the student approval or disapproval of performance, one might investigate what is reinforcing to the student, how best to relay the reinforcement to the student, and observe any behavioral changes which take place. It is the purpose of this particular study to apply learning research to the speech classroom and to determine if positive reinforcement presented by varying modes can result in the improvement of the beginning speaker's speaking skill.

Relevant Research

Since the instructor critique underlies learning in the speech classroom, this review will begin by examining some of the means of critiquing student speeches. Many recommendations have been made concerning the nature of a good critique, but two appear to be directly related to the present consideration: (1) the kind of comments that are made, and (2) the time of the critique in relation to the student's performance.

Favorable, constructive remarks seem to be regarded as the kind of comments composing a good critique.

Dedmon (1967) makes such a recommendation, as do Weaver, Borchers and Smith (1952), Wallace (1954), and Bostrom (1963). However, a descriptive survey of the central United States completed by Lothers (1966) found this recommendation not followed. From the 186 respondents, it was found that 50% give equal weight to favorable and unfavorable remarks; another 25% give greater weight to unfavorable remarks.

In addition to the content of the critique, the time relationship between the critique and the performance is also important. Robinson and Kerikas (1963) suggest that the instructor may offer criticism after every speech, at the end of a series of speeches, and/or during the performance of the speech. They support the latter as the best time for handling errors of delivery. Since this study is concerned with delivery, their recommendation is of particular significance. The Lothers (1966) survey indicates that few instructors actually criticize during the student's speech. The majority of respondents offer their critiques at the end of the speaking day, while only 22 (11.8%) answered that they critique during the speech itself.

It would appear from the preceding discussion that the typical instructor's criticism, the chief classroom tool for skill development, contains as much disapproval as approval and is presented to the student at varying times after he has spoken. The findings of learning research suggest that both practices may be contrary to

conditions of maximum student improvement.

Research stemming from the psychological and educational laboratories shows the consistent finding that responses that are accompanied or followed by immediate reinforcement will be more likely to reoccur within similar situations (Becker, 1963). This finding resembles Thorndike's "law of effect," which states that when "a modifiable connection is made and is accompanied by or followed by a satisfying state of affairs, the strength of the connection is increased; if the connection is made and followed by an annoying state of affairs, its strength is decreased (Hilgard and Bower, 1966). Thorndike later modified the "law" because experimental findings did not show punishment reducing the probability of behavioral repetition (Postman, 1947). However, the experimental findings supported the conclusion that positively reinforced behavior tends to be repeated.

Inherent in the preceding discussion is the relationship between reinforcement and its timing with respect to the behavior being reinforced. Contiguity in time is presupposed as a basis for learning (Shelton, Arndt and Miller, 1961). Students learn best, writes Byers (1963), when they have immediate and valid knowledge of their success or failure. Harms connects the notion of contiguity to the classroom when he writes: "Experimental data indicate that immediately reinforced responses are necessary for efficient learning. Hence, the more responses a student makes which receive immediate reinforcement, the

more rapidly and surely he will learn (1961, p. 216)." Learning, then, may not only be encouraged by the content of the reinforcement but also by the time relationship to the event.

While the acquisition of new behavior requires immediate positive reinforcement, the development of existing behavior also utilizes this kind of reward. Skinner (1953) distinguishes acquisition from development by calling the latter skill. He offers the example of ball throwing where certain responses release the ball at precisely the right instant to achieve maximum distance. Early or late release will decrease the distance that the ball travels and will not be reinforced. Reinforcement which develops skill must be immediate or the precision of the differential effect will be lost.

Delay of reinforcement may be particularly confounding for speech learning. Since speaking is a complex activity, a delay in reinforcement from the deserving act may encourage the repetition of a series of intermediary actions and not the specific action desired (Shelton, Arndt and Miller, 1961). Yet the time for classroom critiquing appears generally to be several minutes after a particular speaking event is completed. Differential positive reinforcement enhances skill development, provided that the reinforcement of the actions occurs immediately after their appearance.

It is assumed that an analogous situation exists between ball throwing and the hand-arm animation of a

speaker. If so, then refinement of the speaker animation requires immediate, positive reinforcement. The delay, as has been noted, caused by waiting until a speaker or series of speakers has finished his or their speeches, may reduce the effectiveness of the reinforcement. Such a reduction can then delay the development of gestural skill refinement. As the Lothers survey indicated, the situation described approximates the speech classroom.

One problem peculiar to the teaching of speech is that the student does not enter the classroom with no experience in the behavior the teacher wishes to encourage. The student has been speaking all his life. Instruction, then, may not require that the instructor encourage the behavior as much as refine the behavior he finds already being emitted by the student. "Good" speech behavior may exist already in the student's repertory of behaviors. What is required of instruction is improvement of existing patterns so they more closely resemble acceptable speaking behavior. This situation resembles Type R conditioning described by Hilgard (1966).

In Type R conditioning, the emitted responses are designated as operants. These operants, once observed, are reinforced. Hilgard writes that "reinforcement cannot follow unless the conditioned response appears; reinforcement is <u>contingent upon response</u> (p. 109)." Skinner relates this operant conditioning to the "law of effect" by saying that "we strengthen an operant in the sense of making a response more probable or, in actual fact, more frequent

(p. 65)." The consequence of the reinforcement of an operant is to increase the rate with which the operant response is emitted. The frequency of response, or emission, becomes a measure of the learning process of the subject (Bigge, 1964). Applied to this particular study, the number of gestures emitted and the increase in animation over time will affect the judges' evaluation of the effectiveness of the speaker's physical behavior.

How often the learner receives reinforcement is also an important consideration. Two general schedules of reinforcement have been studied within the learning laboratory: continuous reinforcement (every or no response reinforced) and intermittent reinforcement (some ratio of responses reinforced. A continuous schedule of reinforcement is the quickest way to increase the frequency of response (Berelson and Steiner, 1964). Sidman indicates this schedule is a powerful technique because the learner will respond even when deprivation is minimal and reinforcement is small (1962).

In addition to educational and psychological research findings, at least one communication concept relates to learning: the concept of feedback. Weiner defines feedback as a "method of controlling a system by reinserting into it the results of its past performance (1954, p. 61)." Feedback, according to Berlo, can be rewarding or nonrewarding to the source of the initial message. In other words, reward (positive feedback) promotes the continuation of current messages, while nonrewarding (negative) feedback promotes

change. The human feedback system is external (1962); that is, the output passes through the environment and returns to the system as input. Feedback, then, resembles reinforcement in the laboratory. Weiner supplies the linkage between feedback and learning. He says that information which proceeds backward from the performance and is able to change the general method and pattern of performance is a process called learning.

Most studies of feedback have used the concept as the independent variable with subsequent changes in verbal learning, delivery and fluency, or some other source behavior attributed to the manipulation of the amount and kind of feedback. Miller, Zavos, Vlandis and Rosenbaum (1961) studied the effect of reward ("good") and nonreward (experimenter silence) feedback on two sequential speakers. While they found that lack of approval of the second speaker did not significantly effect his fluency, ratio of nonfluencies, and his estimate of time spoken, they did find that approval of the first speaker had a disruptive effect upon the performance of the second.

In a later study, Miller (1964) used approval, disapproval and "no comment" as the feedback for student speakers. He found that approval-disapproval variation between two speakers in sequence resulted in a decrease in speaking efficiency in the second speaker. Vlandis (1964) also manipulated reinforcing conditions in the form of feedback. Unlike the Miller experiment, Vlandis used only one speaker and manipulated the feedback during the different

portions of the same speech. The punishment group (received "no" from the experimenter) showed disruptive effects when measured on total utterance and number of nonfluencies. The reward group showed a slight, but insignificant, increase in utterance and similar results for the moments of nonfluency.

Sereno (1964) varied feedback between two sequential speeches for the same person. Using 90 women as subjects and measuring their total verbal utterance and nonfluencies, he found the inferior treatment (less favorable at Time 2 than at Time 1) to produce a significant decrease in verbal utterance at Time 2. Nonfluencies increased at Time 2 under the inferior condition, but did not reach statistical significance. The superior treatment (more favorable at Time 2 than at Time 1) showed an increase in verbal utterance and a decrease in nonfluencies at Time 2, but neither result reached statistical significance.

Blubaugh (1966) studied the effects of positive and negative audience reaction upon fluency and speaker output. Like the above authors, he found greater nonfluencies under negative conditions of feedback than in the positive condition. Rate of speaking decreased and verbal output was lessened in the negative treatment group.

Davis (1967) used two females in a discussion situation with an authority figure (high-low) interjecting "yes" (reward) or "no" (punishment) into their discussion or remaining silent. The reward group had significantly higher total utterance than either of the other conditions, while

the "same" group had a significantly higher total utterance than the punishment group. Nonfluencies were higher for both the disapproval and approval groups.

Like Blubaugh, Karns (1964) used live audiences to manipulate the feedback conditions. One audience was to display an "attentive," "interested," and "accepting posture" of the speaker's message while the experimental audience, on cue, was to show "disagreement" or "lack of understanding." When comparing the recorded speeches with the manuscripts of the prepared speeches, Karns found that changes in the manuscript were preceded by aversive stimuli from the audience.

Amato and Ostermeier (1967) studied the effects of neutral and unfavorable feedback upon the delivery of a student speaker, a study more closely related to this work. Using colored cards to provide feedback, they hypothesized that delivery, measured by observers, would deteriorate in the unfavorable condition. There were six characteristics of delivery which they consolidated into bodily action, eye contact, fluency and nervousness. The hypotheses were supported for all delivery characteristics. The unfavorable feedback condition also prompted a shortening of the speeches by the speakers.

The results of the study were challenged by Combs and Miller (1968) on the grounds that the audience, who administered the feedback, also made the speaker evaluations. Combs and Miller concluded that the speaker differences were due to the response role of the raters and

not the effect of the feedback. To demonstrate their challenge, they conducted a similar experiment with the addition of a favorable condition. This study also measured the audience rating of delivery, content, language and overall effectiveness. The data, which showed a relationship between the feedback condition and student rating, was interpreted as supporting their initial challenge.

In a rejoinder, Ostermeier and Amato (1968) maintain their study supported the original conclusions. The one reason they mention that may deny the response-role challenge is that their subjects were instructed to provide a "majority" of responses in the favorable-unfavorable feedback conditions, rather than to respond only positively or negatively to the speaker. The Combs-Miller study asked the audience-raters to provide only one kind of feedback.

One characteristic common to the mentioned studies is that they were concerned with the deterioration of verbal or physical delivery. The proposed study will focus upon the positive feedback condition to determine if such feedback will increase the desired delivery behavior of the student speakers. The research briefs examined earlier suggest that positive feedback (reinforcement) ought to generate improvement, yet their accompanying experiments irregularly support that conclusion.

Two additional studies need to be mentioned. One, completed under a Works Progress Administration grant, used a signaling device to "feedback" to the student instructor evaluations during the speech (Hayworth, 1939). The device

was an elongated box of six double compartments with an opaque glass front. Speaker-directed comments were printed upon the glass. The speaker was signaled when the light behind the compartment glass flashed, highlighting the printed evaluation remark. This device was used during an entire teaching term. Statistical tests of differences between the experimental and control groups found no significance in aspects of delivery. The results may be due to the directives of the evaluative comments revealed to the speaker when the light flashed. Comments like "more vigor," "walk," "louder," "improve posture," "use gesture," or "extend gesture" all direct the speaker to new courses of action. The implication to the receiving student may be that his current behavior is inadequate. The signals, like other feedback devices mentioned, seemed to be negative in intent.

Another study which used feedback over a term period as a teaching device was done by Dorothy Hinde (1949). She constructed a collection of cards containing comments upon the speaker's delivery and flashed them at the speakers during performance. This lasted for an entire semester with ratings done on the first and last speeches in the course. Ratings were for general effectiveness, bodily communication, use of voice and articulation and pronunciation. Aside from the flash-card group, the other treatment group received comments immediately before the speech, and another received only written criticism after the speech. An overall view showed no significant differences in improvement in the three conditions. All three classes showed improvement in

bodily communication, but it was not significant. All three classes did show significant improvement in articulation and pronunciation. As a sidelight, the students responded adversely to the flash cards. This effect may be due partly to the adverse nature of the comments on the cards. The students also indicated that the appearance of the cards diverted attention and caused embarassment.

Several conclusions may be drawn from the preceding discussion of studies relevant to the intended research: (1) positive reinforcement (approval, positive criticism) of a speaker is recommended but is not always practiced, (2) positive reinforcement (approval, positive feedback, positive criticism) ought to encourage change in behavior but experimental results with people as Ss do not unanimously support the assertion, and (3) research is still needed on the presentation and effects of positive reinforcement within the classroom setting as a means of developing speaking skills in the student of public speaking.

Hypotheses

The major purpose of this study is to determine the feasibility of varying modes of positive reinforcement within the speech classroom with particular attention given to the use of mechanical techniques. The personal-performance skill selected for study is gesture, the movement of the hands and arms. Deutschmann (1964) refers to gesture as one of the four codes found in the communication situation. Eisenson, Auer and Irwin (1963) identify the importance of

gesture to the speaking situation by stating that gestures enhance the speaker's oral symbols, emphasizing meanings and underscoring oral words. For this reason, gesture was chosen as the delivery skill to be used in this study.

One mechanical, electrical device that has been found to be reinforcing is light on-set. Lockard (1963) has reviewed light as applied to rodent behavior and found it a reinforcer. It has been found to be a reinforcer for humans as well. Krasner (1958) identifies light as one experimental cue used in verbal-learning research. Greenspoon (1954) also found light to be reinforcing to human subjects. Cerekwicki and Grant (1967) found light reinforcing to human subjects when conditioning eyelid blinking.

In a study of speaker's attitudes toward topic, audience and self, Huenergardt (1967) used light as a reinforcing instrument. The Ss were instructed that a panel of 20 lights was being controlled by a hidden, listening audience. When a light flashed, a member of the audience was to be regarded as favorably disposed to what was being said. In fact, the E manipulated the number of lights, using 2/20, 6/20, 10/20, 14/20 and 18/20 as conditions of reinforcement. A significant linear trend was found between the percentage of audience approving and the speaker's attitudes. Cahn (1966) used a similar light panel researching attitude change of speakers toward their topics under conditions of positive, negative and no reinforcement. The point here is that light on-set has been used as a positive reinforcer for college students.

In order to accept light on-set as an effective classroom reinforcer, two criteria must be met: (1) light must be found to be reinforcing (able to change behavior), and (2) light ought to encourage the development of gestural skill at least as well as currently used classroom reinforcers. In applying these criteria, this study was placed within a beginning speech classroom, and the effects of light compared with the effects of a mode of reinforcement in current use.

The current mode used by instructors is the written critique. Thus, for this study, two forms of written critique were created, both of which make use of ratings. One form, Minimum Written Reinforcement, relayed to the student a positive comment upon a single aspect of his gesturing. The other written form, Maximum Written Reinforcement, relayed positive comments on a multitude of aspects of student gesturing. The Light On-Set Reinforcement condition flashed a light for each and every gesture found acceptable to the experimental evaluator. This study will compare the impact of the two written modes with that of the light reinforcement condition as measured by the amount of change each condition generates in the student.

Thus, the basic hypothesis under consideration is:

H_1: Minimum Written, Maximum Written and Light On-Set Reinforcement will have significantly different effects in terms of the observed improvement in student animation.

In the above hypothesis, "observed improvement" is the difference between judges' evaluations of student speakers before and after exposure to the reinforcement conditions;

"student animation" is the hand and arm movements displayed during a classroom speech.

Should the results of the study demonstrate that differences have occurred, these differences should appear in a particular order. Since the Maximum Written Reinforcement will show positive evaluations to the student on several aspects of gesturing, it appears that this condition would be providing the student with (1) more information on what elements of gesturing are considered important by the instructor, and (2) more approval of the gestural behavior than would the single comment of the Minimum Written condition.

Therefore, it is hypothesized that:

H_{1a}: Maximum Written Reinforcement student improvement will be greater than student improvement in the Minimum Written Reinforcement condition.

Even though the Maximum Written condition will provide more information to the student than the Minimum Written condition, a similar weakness exists. Both conditions delay the positive reinforcement, and both conditions offer evaluations of the general gestural behavior of the student. The Light On-Set Reinforcement, on the other hand, offers immediate positive reinforcement for a specific act of the hands and arms. This Light On-Set condition most utilizes the theoretic conclusions drawn earlier and so should generate greater change than either written condition:

H_{1b}: Light On-Set Reinforcement student improvement will be greater than student improvement in the Minimum Written Reinforcement condition.

H_{1c}: Light On-Set Reinforcement student improvement will be greater than student improvement in the Maximum Written Reinforcement condition.

If the major and secondary hypotheses of this study can be affirmed by the change (improvement) data, then it may be concluded that immediate, positive reinforcement presented by a light flash, may be a useful tool for speech instructors to use in refining speaking skills of student speakers.

This chapter has provided justification for a research concerned with communications between instructor and student. A brief review of relevant research was presented, as well as a rationale for the study of gesturing as a speaking skill and for the use of light as a reinforcer. One major hypothesis and three secondary hypotheses have been posited to be tested. The procedures used to gather and analyze the data under these hypotheses will be explained in the following chapter.

CHAPTER II

PROCEDURES

The purpose of this chapter is to explain the logistics of this study and the materials and methods used to gather and analyze the data. Methods for the selection of the subjects and for the formation of the measuring instruments are included, as are the descriptions of the several treatments.

Pilot Study

In order to establish procedures for this research, a pilot study was conducted. The subjects were 14 freshmen enrolled in Speech 102, the beginning speech course at Bowling Green State University, during the summer quarter of 1968. This particular class was part of the second of two five-week sessions that comprised the summer quarter.

The course syllabus requested that each student deliver five original speeches, one to be given each day of the week. These speeches could be on any topic of the student's choosing but could not be a "demonstration" speech. Each speech was to be no fewer than two and no more than four minutes in length. Since the summer classes last two hours, the time restrictions permitted every student to speak every day.

The purpose of the pilot study was to determine

(1) if light could be conveniently used within the classroom setting, and (2) if light on-set would be found reinforcing to the students in a classroom. The answer to the first question was found by examining judgments of the instructor-experimenter and a graduate student associate. The answer to the second question was found by measuring changes in the frequency of gesturing from speech 1 to speech 5.

In order to measure changes in frequencies over time, a base-rate for each student's gesturing was established. This base-rate period included the entire speech 1 and the first 1½ minutes of speech 2 during which time no light reinforcement was presented to the student. The associate, acting as an objective counter, watched and counted every gesture (hand and arm movements) emitted by the students. The base-rate ratio of gestures to time was computed by using 15 seconds as one unit of time. Each student had a score which was the average number of gestures emitted for a 15-second time period. This rate of emission was computed for every speech, but only the ratio at speech 1 and speech 5 was used in analysis.

After the base-rate period, the light was flashed for each and every gesture emitted by the student in the remaining speeches. (The associate had been counting similar gestures.) The flash came from a green light bulb of 25-watt power and lasted approximately two-tenths of a second. The bulb sat in a porcelain base placed immediately in front of the speaker's lectern at the front of the room. The operator sat with the counter in the center of the room within easy

view of the lectern.

Each day a different speaking order was followed, and each day the Ss were reminded that the light meant "good gesture." Panels of three students provided free-response comments to the speakers at the end of each speech round. If time permitted, oral comments were made by the instructor, usually dealing with some error common to the majority of the class for a particular speech round.

In order for light reinforcement to be considered effective in the classroom setting, the students exposed should increase the number of gestures emitted during their oral presentations. A perusal of the data collected from the pilot study suggested that no change had taken place in the emission rate of the students. In fact, 10 of the 14 students showed a decrease in the number of gestures exhibited during speech 5 compared with speech 1. Analysis of the data by the sign test (Siegel, 1956) showed this result to reach $p < .090$ level of significance.

Based upon the data, experimenter observations, and discussions with the counter, four recommendations for the main research grew from the pilot exploration:

(1) The light should be shielded from the other class members. When the light was first flashed during speech 2, the nonspeaking class members laughed and whispered. Reactions similar to these occurred for every speaker during that day. This result can be explained within the classical conditioning paradigm in which the negative peer responses may have become associated with the light.

(2) Greater control over the types of speeches should be exercised. While some types of speeches encouraged physical movement (a persuasive speech for civil rights); others did not (a eulogy to Robert and John Kennedy).

(3) More restrictive standards for gestures to be reinforced should be determined. Counting and reinforcing every hand and arm movement was not realistic in a beginning speech class where much of the student's movement springs from fear and nervousness. When such nervous movements were reinforced, the watching students seemed confused about the criteria for evaluation. A "shaping" procedure seemed to be more appropriate for beginning speakers.

(4) A content-related gestural evaluation procedure should be established. Instead of a counter, some instrument apparently needed to be created to permit an instructor evaluation of the movements within the frame of the content and the minimum standards of "appropriate" physical behavior. The statistical results of the study indicated that the number of gestures decreased. However, it was the opinion of the experimenter and the counter that the quality of gesturing (size of movement, relation to content, etc.) did improve during the five speeches.

How these recommendations were incorporated into the main research will become apparent in the remainder of this chapter.

Experimental Conditions

As suggested previously, the intent of this research was to observe and measure differences between classes experiencing dissimilar modes of positive reinforcement. The written mode was chosen because (1) the classroom setting required this study to be as realistic to the standard speech course as possible, and the written critique is used most by the teaching assistants of the course; (2) most speech instructors use written comments to students (Lothers, 1966); (3) speech instructors seem to prefer written remarks as a means of focusing the student's attention upon strengths and weaknesses peculiar to him, and as a means of supplying the student with a permanent record of his own progress (Lothers, 1966).

Besides having current acceptance, the written mode, particularly written scales, has several other qualities recommending its use in this study. The content of the scales, that is, what elements of gesturing are included, can be standardized. This standardization assists in reducing variability which arises from instructor idiosyncrasies concerning gestural elements or from differential nomenclature. Furthermore, phrases expressing the evaluative judgments of the observers can be standardized, which provides a control of the range of written reinforcement.

A written scale has at least two advantages over oral comments (the other mode of instructor criticism most frequently used). The content of the oral comments may interact with verbal and nonverbal codes, strengthening or

weakening the reinforcement, depending upon the perceptions of the student. Such interaction may occur even if the oral comments have been pre-arranged and standardized experimentally (Bostrom, 1963). In addition, public oral comments, if given immediately after the completion of a single speech, may have a disruptive effect upon the speaker(s) following (Miller, 1964).

However, the written mode in the light of learning theory suffers from two distinct disadvantages: (1) reward is delayed, and (2) specific behavior can not be reinforced.

As mentioned previously, light has been found reinforcing to college students. Furthermore, a light flash, because it need not be disruptive to the speaker (as are oral comments), can be used while a student is speaking. Light, therefore, permits the instructor to contiguously reinforce specific speaker behavior. Also, light has an advantage over other mechanical reinforcers in that its apparatus may be relatively simple and easy to transport and assemble. The modes of classroom reinforcement to be compared in this research, then, are written scales and light on-set.

Inherent within each mode are differing amounts of reinforcement. A written scale, for example, may permit several different negative and positive values for each item on the scale. The number of values per item may be varied according to the wishes of the experimenter. The number of values and their placement upon a visual continuum may then appear to be of different amounts of reinforcement. This

perception depends upon the terminal points of the continuum and the placement of values along that continuum. It is assumed, for example, that the value judgment "excellent" would be perceived as a greater amount of positive reinforcement than the evaluation "okay" on the following continuum:

Superior Excellent Very Good Good Okay

By standardizing the number of values available for each and every item on a scale, some control of the perceived amount of reinforcement may be exercised.

In the structuring of the experimental conditions for this study, it was assumed that a scale containing a greater number of items than another will be perceived as offerring more reinforcement. That is, if a scale contains six items of five values each, this appears to be greater than a scale with only one item of five values.

Based upon these apparent differences, a minimum and a maximum written reinforcement condition was selected. Both conditions used the same number of values (a range of four evaluative judgments) and were to be delivered to the student at the same time (immediately upon the conclusion of his speech). They differed in the number of items on the scale presented to the student. The minimum written scale consisted of only one item, while the maximum written scale included nine items.

The choice to use only one scale in the minimum written condition was based on the observation that most instructors' critiques often contain only one or two remarks

on several variables found in the speaker's performance. Time, of course, forces the critiquing instructor to limit the characteristics he wishes to emphasize to the student and the number of comments he can make concerning those few characteristics.

How the items of the scales were chosen and the entire instrument was distributed will be discussed later.

In the light on-set condition, the value system was dichotomous --- present or absent. However, since light could be flashed for each and every desired gesture, the amount of reinforcement was related to the number of gestures emitted by the speaker. The more the speaker movement was deemed acceptable by the evaluator, the greater was his reinforcement. It is this characteristic that could make the light on-set different from either the written conditions: reinforcement could be presented as often as the speaker's behavior met certain standards. The written conditions, however, could be presented to the speaker only once, and then, some time after the desired behavior had been emitted.

The characteristics discussed above operated in this research as three reinforcement conditions: Minimum Written Reinforcement (MNW), Maximum Written Reinforcement (MXW), and Light On-Set Reinforcement (LOR). Figure 1, page 27, provides a brief comparison between the conditions:

Written		Light
Minimum	Maximum	
1-item scale	9-item scale	Flashed for all gestures meeting minimum standards of "acceptability"
4 positive values	4 positive values	2 values: present and absent
Given Ss immediately after his speech	Given Ss immediately after his speech	Given Ss immediately after gesture emitted
Specific evaluation referring to the general gestural behavior of the Ss	Specific evaluation referring to the general gestural behavior of the Ss	General evaluation of specific gestural behavior of the Ss

Figure 1. Description of Experimental Conditions

As can be seen in Figure 1, each condition was an aggregate of several variables: time, amount and kind of reinforcement.

Experimental Evaluators

The experimental evaluators (EE) who administered the reinforcing conditions were doctoral students in the Department of Speech at Bowling Green State University. All had taught at least four years before beginning their graduate studies. Several discussions took place during which time the purpose of the research was explained and the EE's place within the research clarified. Each EE worked with the reinforcement instrument prior to the research. During one such training session, two student speakers were invited to prepare and give speeches to be witnessed by the EE. Each EE completed his written ratings (or flashed the light) for each speaker. These evaluations were then discussed

and any deviations particularly analyzed. The training session continued until each EE felt confident in the use of his instrument and in the similarity of his evaluations with the other two EE.

Subjects

Subjects (Ss) were undergraduate students enrolled in the beginning speech course at Bowling Green State University during the spring quarter of 1969. The course structure was divided between a lecture and two drill sections per week. Nine drill sections with 104 students registered were chosen for the experiment.

Of those students who originally registered, 16 missed one or more of the first seven drill section class periods and were disqualified as Ss. These students were not informed that they had been disqualified, thus they continued to complete the course requirements. The 88 students who completed the experiment and who were eventually included in the analysis were divided between the experimental conditions as follows:

Table 1. Distribution of Subjects by Treatment

Treatment	Enrollment	Disqualified	Total
Minimum Written	35	6	29
Maximum Written	34	3	31
Light On-Set	35	7	28
Total	104	16	88

There appeared to be no systematic pattern in the Ss disqualification. Table 2 identifies the times at which students were disqualified as Ss.

Table 2. Subject Disqualification by Speech Assignment

	pre	1	2	3	4	post	Total
MNW	2		1	1		2	6
MXW			1		1	1	3
LOR			4	2		1	7
Total	2	0	6	3	1	4	16

Several of the Ss missed more than one speaking assignment. However, the first time they were absent constituted the disqualification time. While speech 2, the fourth class day and second treatment period, shows the greatest loss, there appears to be no systematic loss within treatments or across speeches.

Measuring Instruments

To determine the effects of differing modes of positive reinforcement upon the gesturing of student speakers, an instrument for recording judges' evaluations was created. In creating the measuring instrument, a free-response questionnaire was sent to members of the teaching faculty and graduate students in the Department of Speech at Bowling Green State University (see Appendix A, page 78). Each was asked to list characteristics of good gesturing which he sought to teach in the beginning speech course. From the

responses given, 14 characteristics were identified. These 14 were then randomly listed in a second questionnaire sent to the same personnel (see Appendix A, page 79). The teachers were asked if there appeared to be overlapping characteristics in the listing. If so, they were to identify the characteristics that seemed similar. Also, they were asked if any important characteristics did not appear in the listing. From these reactions, a total of nine characteristics of good gesturing were found. The characteristics identified were: "appropriate to content," "clarity," "completeness," "coordinated with other bodily action," "definiteness," "directed toward the audience," "fitting the speaker," "spontaneity" and "visibility." The nine characteristics were used to form part of the total measuring instrument. They were placed on a scale resembling the semantic differential (seven-step). The extreme positive position was assigned the value "7," while the extreme negative position received the value "1." The instrument was assumed to be unidimensional (see Tables 3 and 4, pages 43 and 44), allowing the combined judge ratings for all nine items to constitute a student's Gestural Evaluation Score (GES).

Given that gesturing bears relationship to the total physical behavior of a speaker (posture, walking, etc.), then improvement in gesturing ought to improve the impressions an evaluator might have of the total physical behavior. To test this improvement, an additional measure was created. Five sets of paired adjectives were selected from the Berlo, Lemert, Mertz (1961) research on credibility.

The adjectives chosen from the Dynamism factor were: "bold-timid," "aggressive-meek," "energetic-tired," "confident-unsure" and "colorful-dull." The semantic differential format was used to present the Dynamism items to the judges. By adding together the judges' evaluations for each item and adding the items together, a total Physical Performance Score (PPS) was computed and analyzed. A sample of the instrument is contained in Appendix B.

Since the speaker evaluations were made by panels of three judges each, a GES for a single Ss could range from a low of "27" (all judges awarding "1" for all nine scale items) to a high of "189" (all judges awarding "7" for all nine items). The PPS could range from a low of "15" to a high of "105."

Logistics

The characteristics of good gesturing found by polling the teaching faculty were used not only in the measuring instrument. The nine elements formed the content of the item scales used in both written reinforcement conditions. However, the number of items (as mentioned earlier) differed between the two conditions.

Minimum Written Reinforcement (MNW) Condition

In the MNW condition, the characteristics of good gesturing were divided into nine separate scales. The separate scales were attached to the inside of a standard 9-by-12-inch file folder. When the folder lay open, all scales were clearly visible to the EE. As he watched the speaker, he decided which characteristic of good gesturing was best

performed. After deciding which characteristic he wished to reinforce, the EE then circled the evaluative comment (of the four available) which most clearly expressed his judgment of that particular characteristic. When the student had finished speaking, but before he had received the instructor's written comments, he was given the gestural scale completed by the EE. A sample minimum reinforcement instrument can be found in Appendix C, page 82.

For each class he evaluated, the EE had a folder containing the nine scales and a class roster.

Maximum Written Reinforcement (MXW) Condition

In the MXW condition, the characteristics of good gesturing were combined into a single evaluation sheet of nine scales. The evaluation sheets were attached to the inside of a standard 9-by-12-inch file folder. As the EE watched the speaker, he judged him on all nine characteristics. The EE circled the evaluative comment which expressed his judgment of the speaker's performance of each and every characteristic. A sample maximum reinforcement instrument appears in Appendix D, page 85.

For the three classes he evaluated, the EE had a folder containing the scales and a class roster.

Light On-Set Reinforcement (LOR) Condition

The task of the EE for the LOR condition differed from the tasks of either EE for the written conditions. Whereas the EE for both written conditions made evaluations of specific gestural characteristics for a completed speech, the EE for the LOR made an evaluation of each and every hand

and arm movement of the speaker within a single speech. The nine items of good gesturing were applied to each gesture, an evaluation made, and the light was flashed or not. If the gesture was considered "acceptable," the light flashed; if the gesture was considered "unacceptable," the light was not flashed.

The light apparatus was 18 inches wide, 11 inches high and 12½ inches deep. The base consisted of a piece of 2-by-6-inch lumber chosen to ensure a solid mounting for the light fixture and to give stability to the whole apparatus. The sides, top and back were made of cardboard. When resting upon a desk, the apparatus appeared to be a cardboard box with one long side missing. The light was mounted on the board in the center of the box. The fixture itself was a lightbulb encased in a green glass cone resting in a black metal base. A 1½-by-2½-inch opening covered with green gelatin was located two inches from the rear-center of the top of the box. This arrangement enabled the EE to see the light as it flashed, yet did not permit light to escape.

The power for the light was a single six-volt battery encased in a box 6-by-6½-by-3 inches. To flash the light, the EE pushed a button on top of the battery casing. While the bulb was small and powered by a mere six volts, its flash produced 25 candlepower, sufficient light to be seen from several feet. However, the bulb's power was not enough to produce light reflections off objects in the room, which might distract the audience or cause them to react. As a further precaution, the EE sat closest to the speaker's

lectern, alone in the front row.

The EE for the LOR also had a 9-by-12-inch file folder for each class he worked. It contained a class roster and a listing of the nine characteristics of good gesturing used in the study.

Procedures for Data Collection

In order to minimize classroom differences, only three drill instructors were chosen. They were selected with the assistance and recommendation of the coordinators of the beginning course. All three had been undergraduate speech majors; all were graduate assistants majoring in the area of Rhetoric and Public Address. Each instructor taught three drill sections, the normal graduate teaching load during the spring quarter. To further control for instructor variability, the three experimental conditions were randomly assigned to the sections of each instructor. Thus, every instructor had one section in each of the three conditions. To prepare them for the research and their place in it, the three instructors met with the researcher in an extended session prior to the beginning of the spring quarter. Individual discussions also were held prior to and during the running of the experiment.

Classroom Assignments

Before the start of the spring quarter, the drill instructors and the experimenter structured the course. It was decided that the first seven drill sections would be devoted to the administration and execution of the experiment. The intent of the common syllabus was to discourage variability

due to differences in speaking assignments. A complete assignment sheet appears in Appendix E, pages 86-88.

During the first class period of the quarter, the procedural matters of the course and the experiment were clarified. Each student received a copy of the syllabus, and each assignment, including those of the experiment, was explained, and questions were answered. Grading and attendance were also discussed, and the class roster corrected. It was explained that grades would be given by the instructor, but that the Ss would not receive them until after the first six speeches were completed. The Ss were told that two grades would be recorded --- the highest grade given and an average of the other five speech grades. The Ss were told that they would be treated differently from other sections of the course, but that such "experimentation" was common to a multi-sectioned course like Speech 102.

Administration of the Pre-Test

Sixteen faculty and teaching graduate assistants were randomly assigned into nine panels of three judges each. While several judges evaluated together on more than one panel, no two panels were identical. They were instructed to evaluate the hand and arm movements of the students, using the first nine scales of the evaluation sheet. The final five Dynamism scales were to be used to indicate their impressions of the general physical behavior of the speakers. Posture, facial expression and walking were to be evaluated by these scales. These instructions and a copy of the evaluation sheet were given to the judges prior to the time that they

were to hear speeches.

The students spoke in alphabetical order. The instructor called each student by name and spelled the last name of the student for the benefit of the judges. Such a procedure kept the instructor in control of the class and allowed him to set the pace of the speaking order. Since the instructors were also evaluating, critiquing and grading, it was necessary to permit them to determine the amount of time needed between speakers.

After the Ss' name was called, he delivered his speech from the front of the classroom. Upon completion of the speech, the Ss received his written critique from the drill instructor. The panel of judges independently marked the evaluation sheet but did not inform the Ss, the instructor, or the other judges of their evaluation. When forms and critiques were given to the Ss, the instructor called for the next Ss, and the process was repeated until all Ss had spoken. If time permitted, oral remarks were made by the instructor on any aspect of speaking except that under investigation.

Administration of the Reinforcement Treatments

During the next four drill section periods, other speaking assignments were performed, and the reinforcement treatments administered. The instructor called the Ss and spelled the last name for the benefit of the experimenters. While an alphabetical order was maintained, each day the instructor began at a new, predetermined point in the alphabet.

At the conclusion of a speech, the Ss received a

written critique from the instructor and a completed evaluation sheet from the EE. The material from the EE preceded the material from the instructor. Once the Ss had received his material, the next speaker was called. If time permitted, an oral critique by the instructor was given. The EE said nothing to the students or the instructor.

In the LOR group, the procedures were the same except that the Ss were told the light meant "good gesture" (see Appendix F, page 89). Only the instructor wrote comments to the Ss. The EE in the LOR group sat in the first row in front of the lectern. The light mechanism was placed before him so he could clearly see the Ss and the light in the box. A light flash lasted approximately two-tenths of a second --- the time needed to press and release the button.

Administration of the Post-Test

Seventeen judges were randomly assigned into nine panels of three judges each. No panels were the same, nor were any post-test panels identical to any pre-test panels. Twelve judges evaluated during both time periods, but only two heard in the post-test students from classes they had heard in the pre-test. The same evaluation sheet was used, leading to a second GES and PPS for each student.

Classroom procedures used in the preceding five speaking periods remained constant for the post-test. As in the earlier sections, the instructors were not informed of the judges' evaluations. As each panel returned its evaluations, the experimenter checked them for errors. Fortunately,

no judge's errors caused a Ss to be disqualified from analysis.

Upon completion of the experiment, the drill sections returned to more normal procedures. That is, fewer speeches were given, and persons from outside the class were not present.

After the judges had completed the post-reinforcement ratings, the differences between the evaluations of the pre- and post-test speeches were computed. These difference scores constituted the basic data for this study.

Statistical Design

Since the purpose of positive reinforcement is to improve the gestural skills of the students receiving the reinforcement, it was decided to use improvement scores as the basic data for analysis. That is, the improvement score was the difference for each student who participated in the research as shown by changes in both the GES and PPS from Time 1 (T_1) to Time 2 (T_2). The relative effect of the modes of reinforcement presented to the students would be reflected in variations of these differences.

Using intact educational groups (speech classes) for Ss, there was a possibility that significant differences in gesturing ability existed before the reinforcement conditions were applied. To detect such differences, a one-way analysis of variance was computed on the pre-test GES and PPS evaluations. If a significant F-value was found, a single-factor covariance design was to be used in which the pre-test GES and PPS for each Ss would be the covariant for

analysis of the post-test GES and PPS. On the other hand, if no significant F-value was found for the pre-test data, the basic statistical design was to be a one-way analysis of variance, using difference (improvement) scores as the dependent variable for both the GES and PPS. Should the analysis of the improvement scores (by either covariance or one-way analysis of variance) have resulted in a significant F-value, then t-tests were to be applied to make individual comparisons between groups.

The above statistical design was to hold if the data did not violate the mathematical assumptions of the parametric analysis of variance routines. If, however, those data did not satisfy the assumptions, then nonparametric techniques would be used in analysis. Specifically, the Kruskal-Wallis one-way analysis of variance would be applied to the difference scores. If a significant difference was found between the three modes of reinforcement, the Mann-Whitney U test would be computed to make individual comparisons between the conditions. These two statistics approximate the parametric tests described above.

Regardless of the statistical procedures employed in analysis, the major null hypothesis for the study of the GES may be stated:

$$H_{O_G}: \mu_{MNW} = \mu_{MXW} = \mu_{LOR}$$

The major null hypothesis for the PPS, using improvement score for analysis, is:

$$H_{O_P}: \mu_{MNW} = \mu_{MXW} = \mu_{LOR}$$

If the major null hypotheses were rejected, then further

statistical analysis would test the following hypotheses and their alternatives regarding the direction of the differences for both the GES and PPS evaluations:

H_{O_a}: $\mu_{MNW} = \mu_{MXW}$ $\quad$ H_{1_a}: $\mu_{MXW} < \mu_{MNW}$

H_{O_b}: $\mu_{MNW} = \mu_{LOR}$ $\quad$ H_{1_b}: $\mu_{LOR} < \mu_{MNW}$

H_{O_c}: $\mu_{LOR} = \mu_{MXW}$ $\quad$ H_{1_c}: $\mu_{LOR} < \mu_{MXW}$

Summary

This chapter has explained the pilot study, the selection of subjects and the way in which the instruments for reinforcement and measurement were created. The speech class course structure was explained, as well as the procedures for gathering and analyzing the collected data. The results of the statistical comparison and an explanation of the results of the experiment are reported in the next chapter.

CHAPTER III

RESULTS

The analyses reported in this chapter were conducted to test the hypotheses (expressed previously) concerning differences between classes in a beginning public speaking course after each had been exposed to varying experimental conditions. The object of the statistical tests was (1) to determine if there were differences between the treatment groups, and (2) to determine (by cellular comparisons) where those differences occurred. As preliminaries to the actual testing of the hypotheses, several questions had to be answered: Was there justification for adding across items of the measuring instruments to produce a total GES and PPS score? Did the judging panels reach some minimum level of rater reliability? Have the assumptions underlying the F- and t-tests been satisfied to warrant the application of these two statistics? Since the eventual analyses of results depend upon satisfactory answers to these questions, they will be investigated first before proceeding to the testing of the major hypotheses of the research.

Preliminary Analysis

The individual items of the GES scale must bear some positive relationship in order to add across the items and arrive at a single total score. Using the judges' responses

on the nine gesturing items from the pre-test, product-moment correlations were computed between each item and every other item and between each item and the summated score. The correlation between the separate items and between the items and the total score indicate the assumption of unidimensionality to be warranted. Table 3 (see page 43) contains the intercorrelation matrix for the GES items and the GES total score at T_1. The correlations are judged sufficiently high to conclude that the items share a positive relationship.[1]

Since several judges evaluated only at T_2, correlations were computed for the GES items and summated at that time as well. It seemed possible that the additional judges might not use the instruments as those who had already evaluated speeches. If so, their inexperience might change the relationships between the items. The intercorrelation matrix for the GES at T_2 appears in Table 4, page 44. Once again, the high correlations suggest the assumption of unidimensionality to be merited. Apparently, judges who rated only at T_2 did not negatively influence the scale.

In using the PPS as a measure of general physical behavior, the assumption of unidimensionality was made again. The correlations reported in Table 5, page 46, support the assumption that the PPS also is a unidimensional measure of

---

[1]While additional statistical procedures are available for testing unidimensionality (factor analysis), because of the small number of evaluations given plus the confounding influence of repeated measures, it was decided to support the assumption with the correlation matrix alone (Guildford, 1954).

Table 3. Correlations Between Items for the Gestural Evaluation Score, Time 1

	Vis	Cla	Def	Com	Nat	Spo	App	Coo	Dir	Tot
Visible (Vis)	---									
Clarity (Cla)	831	---								
Definite (Def)	797	862	---							
Complete (Com)	777	911	858	---						
Natural (Nat)	809	764	683	747	---					
Spontaneous (Spo)	803	843	809	885	802	---				
Appropriate to the Content (App)	803	849	770	852	857	829	---			
Coordinated with Bodily Action (Coo)	829	849	784	842	884	839	910	---		
Directed toward the Audience (Dir)	761	853	796	888	772	840	893	890	---	
Total (Tot)	899	934	885	932	890	923	934	948	928	---

Table 4. Correlations Between Items for the Gestural Evaluation Score, Time 2

	Vis	Cla	Def	Com	Nat	Spo	App	Coo	Dir	Tot
Visible (Vis)	---									
Clarity (Cla)	899	---								
Definite (Def)	890	941	---							
Complete (Com)	886	936	931	---						
Natural (Nat)	811	819	799	843	---					
Spontaneous (Spo)	894	934	938	949	858	---				
Appropriate to the Content (App)	800	809	814	819	925	829	---			
Coordinated with Bodily Action (Coo)	697	752	764	801	820	804	831	---		
Directed toward the Audience (Dir)	823	855	879	885	865	918	844	780	---	
Total (Tot)	921	948	949	960	924	969	915	862	936	---

gestural behavior. Table 5 (see page 46) reports the intercorrelations for both T_1 and T_2 and, at both times, the correlations seem sufficiently high. The decision to add across items of both measures in order to use total GES and PPS scores was sustained by the correlation data.

As reported earlier, it was assumed that hand and arm movements do bear some relationship to the general physical quality of a speaker. Given that assumption, the PPS items were intended as a second measure of gestural performance. However, for the PPS to be a reliable measure (for the assumption to be supported) required that the items of the PPS show positive correlations to the items and total score of the GES. The results of product-moment correlation computations for both time periods are reported in Table 6, page 47. The figures on the left report the relationship of the GES items at T_1 to the PPS items at T_1; the figures on the right report the correlations between the GES and PPS items at T_2. While the correlations are not as high as those computed for the GES and PPS as separate units, the two measures appear sufficiently related to warrant using them as two measures of related, but not necessarily the same, characteristics. It should be noted that the total GES correlated with the total PPS at $T_1 = .840$ and $T_2 = .803$, accounting for over 60% of the variability. The decision to use both the GES and PPS as measures of gestural performances appeared reasonable in light of the correlational data reported.

The next question raised concerned the reliability

Table 5. Correlations Between Items for the Physical Performance Score, Time 1 and 2

	Bol	Agg	Ene	Con	Col	T-1	Bol	Agg	Ene	Con	Col	T-2
Bold (Bol)	---											
Aggressive (Agg)	949	---										
Energy (Ene)	819	840	---									
Confident (Con)	864	848	731	---								
Colorful (Col)	895	880	837	835	---							
Total - 1 (T-1)	961	962	902	910	951	---						
Bold (Bol)							---					
Aggressive (Agg)							971	---				
Energy (Ene)							863	880	---			
Confident (Con)							854	842	820	---		
Colorful (Col)							820	817	878	909	---	
Total - 1 (T-1)							955	955	938	939	937	---

Table 6. Correlations Between Items for the Gestural Evaluation Score and Physical Performance Score, Time 1 and 2

	Bol	Agg	Ene	Con	Col	PPS T-1	Bol	Agg	Ene	Con	Col	PPS T-2
Visible	707	693	722	520	727	719	703	741	733	594	663	726
Clarity	773	748	692	580	722	749	751	785	764	695	738	790
Definite	721	735	698	560	663	717	792	812	789	717	747	816
Complete	778	763	692	643	706	763	775	810	781	740	755	817
Natural	792	739	798	642	805	807	671	672	720	697	722	737
Spontaneous	786	769	706	606	743	769	779	800	783	727	767	816
Appropriate to the Content	801	771	754	674	794	810	602	601	658	597	662	660
Coordinated with Other Bodily Action	809	776	777	692	836	831	546	555	566	621	653	623
Directed toward the Audience	793	752	706	654	724	774	713	729	709	685	720	753
GES Total	841	814	793	673	815	840	756	776	777	723	766	803

of the judging panels. In other words, did the raters reach a sufficient level of reliability to warrant the addition of the judges' scores for each item? Guilford (1954) recommends the use of intraclass correlations as a method for determining reliability among raters. Essentially, the formula below provides for an average intracorrelation:

$$r_{kk} = \frac{V_p - V_e}{V_p}$$

where r_{kk} = variance for the people rated

V_p = variance for the people rated

V_e = variance for error

The r_{kk} was computed for each panel of three judges for each item on both instruments for T_1 and T_2. The median reliability at T_1 and T_2 for each instrument item of the GES is reported in Table 7, page 49.

The rater reliabilities for the PPS are reported in Table 8, page 49.

The r_{kk} computations for each panel used in the research and for each item in the GES and PPS scales are collected in Appendix G, pages 92-94. While there is expected variability between the panels, the reliabilities appear high enough to warrant adding across judges of an evaluating panel in order to arrive at a single score. Furthermore, examination of panel reliabilities shows no treatment group systematically affected by rater variabilities.

Having examined the a priori assumptions of this research and found reasonable justification for using two measures (adding across items and judges), the next preliminary analysis tests the underlying assumptions of the F- and

Table 7. Median Rater Reliabilities for the Gestural Evaluation Score, Time 1 and 2

Items	Time 1	Time 2
Visible	801	700
Distinct	742	714
Definite	668	648
Complete	586	753
Natural	760	557
Free	660	826
Appropriate to the Content	606	576
Coordinated with Other Bodily Action	639	671
Directed toward the Audience	663	714
Total	836	678

Table 8. Median Rater Reliabilities for the Physical Performance Score, Time 1 and 2

Items	Time 1	Time 2
Bold	770	747
Aggressive	639	763
Energy	613	673
Confident	616	742
Colorful	766	814
Total	517	779

t-tests. Both tests make the assumption of a normal distribution. Examination of the raw score data at T_1 shows a positively skewed distribution for each reinforcement condition. Winer (1962) writes that the F- and t-tests are robust regarding the normality assumption, even when the distribution is "markedly skewed (p. 34)." Lindquist concurs, suggesting that, for the F-value, the distribution is "insensitive to the form of the distribution (1956, p. 86)." Given these comments, the skewed form of the distribution seems insufficient reason to reject the parametric tests.

Both tests also assume homogeneity of variance. As noted above, the F- and t-tests seem relatively uninfluenced by failure to satisfy this assumption. Moderate amounts of deviation may be accounted for by setting more rigid limits for rejection, according to Lindquist (1956). However, examination of the data suggests that heterogeneity in this study might be extreme. This deviation was detected in the standard deviations computed for the total GES and total PPS scores at T_1. Since these were the scores to be eventually used in the analysis, it seemed necessary to determine if the assumption of homogeneity was violated. Bartlett's Test for Homogeneity of Variance, as described in McNemar (1962), was performed. This test is appropriate for samples of unequal sizes. The results for the GES are reported in Table 9, page 51. With such variability, the mathematical assumption underlying the F and t routines appears to be violated.

Table 9. Bartlett's Test for Homogeneity of Variance on Gestural Evaluation Score, Time 1

	S.D.	$S.D.^2$	n - 1
MNW	24.24	587.5776	28
MXW	43.42	1885.2964	30
LOR	22.82	520.7524	27
X^2	15.3133 with df = 2, $p < .001$		

While the initial observation of the PPS did not reveal such extreme variability as appeared in the GES, differences did appear, and so these differences were also tested. The PPS variability (reported in Table 10) approaches statistical significance and so recommends that analysis of variance might be suspect if used for analysis.

Table 10. Bartlett's Test for Homogeneity of Variance on Physical Performance Score, Time 1

	S.D.	$S.D.^2$	n - 1
MNW	14.73	216.9729	28
MXW	20.60	424.3600	30
LOR	14.29	204.2041	27
X^2	4.9505 with df = 2, $p < .10$ but $> .05$		

Kerlinger (1965) and McNemar (1962) both argue to use parametric statistics because they are more sensitive, and so, more powerful than are nonparametric statistics.

However, both also admit that serious violations of the assumptions underlying the F- and t-tests may justify turning to the distribution-free statistics.

In view of the departures from the assumption of homogeneous variances between the groups comprising the experimental conditions, coupled with the deviation from normality of the distributions, the decision was made to turn from the F- and t-tests to applicable distribution-free statistics.

The test chosen to replace the parametric F was the Kruskal-Wallis one-way analysis of variance (H statistic) described by Siegel (1956). This statistic tests the hypothesis that k independent samples had been drawn from the same population, like the parametric F, but it uses ranks as the variable. The H test assumes that the variable under study has a continuous distribution, an assumption apparently met by speaker animation. A second assumption requires ordinal measurement which seems satisfied by the semantic differential format used to collect the judges' evaluations. The Kruskal-Wallis test, writes Siegel, "seems to be the most efficient of the nonparametric tests for k samples (p. 194)." When, in fact, the assumptions of the F-test are met by the data and the H test is used, the power efficiency of the H statistic is 95.5% when compared to the F-test. This statistical test, applicable to groups of unequal size, appears to be the most appropriate substitute for the parametric F.

Should difference be found in the improvement score analysis, the Mann-Whitney U test, which also uses ranks as

the variable, would be computed to make individual comparisons between reinforcement conditions. This statistical procedure for data analysis roughly approximates the parametric t-test when further analysis of a significant F is required.

The final preliminary analysis was to test for group difference at T_1. In order to minimize the possibility that differences at T_2 result from initial differences between the groups and were not the result of the experimental experiences, the H test was computed for T_1. With the level of rejection .05 and $df_2 = 2$, the chi-square distribution requires 5.99 to be significant.

Table 11. Kruskal-Wallis Test on Gestural Evaluation Score, Time 1

	Sum of Ranks	n_j
MNW	1444.5	29
MXW	1248.5	31
LOR	1222.5	28
H	1.6811 with df = 2, and α = .05 n.s.	

No such level is reached by the data reported in Table 11. The H statistic, however, is influenced by ties in the ranks. Of the 88 rankings, 51 (57.95%) were ties. The recommended correction for tied ranks was computed, using the formula

$$H_{corr} = \frac{H}{1 - \frac{T}{N^3 - N}}$$

where $T = t^3 - t$; t is the number of tied observations in a group of tied scores

The result of the computation was that $H_{corr} = 1.684$, which, with df = 2 and the significance level = .05, was not significant.

Similar computation was made for the PPS at T_1, with the conclusion that no significant differences existed between the groups at that time.

Table 12. Kruskal-Wallis Test on Physical Performance Score, Time 1

	Sum of Ranks	n_j	H	df	α	X^2	p
MNW	1526.5	29	5.033	2	.05	5.99	n.s.
MXW	1355	31					
LCR	1034.5	28					

However, with the H approaching a significant chi-square and the H usually increased when corrected, the correction for ties was computed. The resulting H increased to $H_{corr} = 5.038$, which was not significant. From the results of testing at T_1 for the GES and PPS, it seemed reasonable to conclude that the groups were drawn from the same populations, and that any differences which might appear at T_2 would be a result of the experimental conditions. Therefore, change score data may be used to test the research hypothesis.

Testing the Research Hypothesis

Since improvement was hypothesized for all groups at T_2, the nonparametric H was applied to difference scores

computed for each Ss by subtracting his evaluation at T_2 from his judges' evaluations at T_1. The differences for each Ss were then rank-ordered for the combined groups. Those ranks were used in the H statistical computation with the results reported in Table 13. While the H does not reach significance, the value falls between a chi-square of 3.22 ($p = .20$) and 4.60 ($p = .10$).

Table 13. Kruskal-Wallis Test on Gestural Evaluation Score Change Scores

	Sum of Ranks	n_j	H	df	α	X^2	p
MNW	1054	29	4.163	2	.05	5.99	n.s.
MXW	1467	31					
LOR	1395	28					

Forty of the scores (45.45%) were involved in tied ranks, so the correction was computed. The value of H increased to $H_{corr} = 4.164$, which did not reach the necessary $X^2 = 5.99$ to be considered statistically significant. With no significance in the GES, it was suspected that the PPS would also show nonsignificance. The results of the Kruskal-Wallis test for the PPS are reported in Table 14.

Table 14. Kruskal-Wallis Test on Physical Performance Score Change Scores

	Sum of Ranks	n_j	H	df	α	X^2	p
MNW	1109	29	3.007	2	.05	5.99	n.s.
MXW	1387.5	31					
LOR	1419.5	28					

With 50 of the 88 ranks tied, the correction for ties was computed. The H_{corr} for the PPS became 3.0089, which fell between 2.41 (p = .30) and 3.22 (p = .20), but was not significant.

Despite the fact that neither the GES or PPS reached statistical significance, apparently changes were taking place within the treatment groups. For both dependent measures, but particularly the GES, the data tends toward the predicted direction. Given that tendency, it was decided to look at the data for evidence of change.

Trends in the Data

In Table 15, page 57, the means and medians, by experimental condition, are reported. The differences that appear between these two measures of central tendency may be due to a few extremely high scores which appear in each group and "pull" the means upward. For example, with a potential GES of 189 at T_1, three Ss in the MXW received scores higher than 155, and four other Ss received scores of 101 or higher. In the MNW, two Ss scored over 120, while in the LOR, two Ss scored over 100. At T_2, the total number of scores over 100 was 13: two in the MNW, six in the MXW and five in the LOR. The point to be made is that the median may be more representative of the groups' scores than the mean. It is to be noted, however, that both measures change between the two experimental times.

It is apparent from Table 15, page 57, that changes were taking place in the various groups. At T_1, for both evaluation scores, the means and medians of the MNW were

larger than those of either of the other groups. However, at T_2, the MNW means and medians are lower than either group in the GES and are lower than the LOR in the PPS. The displacement in central tendency measures, particularly in the GES, may be due to the limited amount of reinforcement given the Ss in the MNW condition. That is, one evaluative comment may not relay sufficient, useful information to the beginning speech student for him to be able to improve his gestural behavior. Furthermore, the data seem to indicate that the MNW condition differed from the MXW condition, supporting (post hoc) a pre-research assumption.

Table 15. Means and Medians by Conditions, Time 1 and 2

	GES			PPS	
	1	2		1	2
MNW	64.03	53.79	mean	48.55	44.76
	60	47	median	46	43
MXW	63.39	71.81	mean	44.84	44.65
	54	58	median	40	41
LOR	58.39	71.57	mean	38.71	45.39
	51	67 (computed)	median	35 (computed	41

But, a countering increase in the evaluations of the MXW and LOR was also occurring. In fact, the evaluations of 87 Ss changed between T_1 and T_2 : 43 Ss showed positive change, and 44 Ss showed negative change. The description

of the Ss movements between the two time periods appear in Table 16. The nonparametric sign test (Siegel, 1956) was applied to the data. It is to be noted that the direction of change in the MNW is in the wrong direction. Sign test analysis shows statistical significance for the number of Ss moving in the negative direction. The direction of movements of the groups (seen in Table 16) seems to relate to the magnitude of the changes reported in Table 15, page 57.

Table 16. Number and Direction of Evaluative Changes by Conditions, Gestural Evaluation Score

	MNW		MXW		LOR	
	+	-	+	-	+	-
	9	19	18	13	16	12
Sign Test p	.0445		.2358		.2877	

However, despite these apparent changes, no statistical support for the hypothesis that differences will appear between the experimental groups after exposure to the treatments has been found. On the other hand, it is not possible to conclude that the treatments had a random effect. Tendencies appear, particularly in MNW, where the direction of change was opposed to that predicted, which seems to indicate the experimental modes of reinforcement had a selective impact upon the Ss.

Since no significant differences appeared between conditions, it was decided to look within groups to see if changes had taken place. The information presented in Tables

15 and 16, pages 57 and 58, would encourage the suspicion that the various modes of reinforcement did not improve the gestural performance of the Ss. To test that suspicion, the nonparametric Wilcoxon Match-Pairs Signed-Ranks described by Siegel (1956) was computed for each group. With large samples (N = 25), the Wilcoxon approximates the Z-distribution with the formula

$$Z = \frac{T - \frac{N(N + 1)}{4}}{\sqrt{\frac{N(N + 1)(2N = 1)}{24}}}$$

where T = smaller sum of like-signed ranks
N = number of pairs less those which show d = 0

When compared with the t-test, the Wilcoxon has the efficiency of 95.5%. In Table 17 the results are reported for both the GES and PPS of the MNW condition. As already noted, the changes

Table 17. Wilcoxon Z-Scores for Gestural Evaluation Score and Physical Performance Score, Minimum Written Reinforcement

	GES	PPS
T (9+)*	146	(11+)* 143.5
N	28	29
Z	1.29	1.60
p	.0985 n.s.	.0548 n.s.

*Smaller sum of like-sized ranks

in the MNW were in the nonpredicted direction for both measurements. While there are no significant differences within the MNW between T_1 and T_2, the probabilities reported in Table 17 approach the alpha probability of .05.

The negative results found in the MNW suggest that

the MNW condition seems unable to encourage positive change in the gestural behavioral of the Ss. In this condition, the Ss received a single gestural comment supplemented by the instructor's critique of other physical and contextual matters. This experimental condition seemed most like the speech critique of the classroom where only a few comments may be made to the student in any one critique. Since the majority of students in each class changed in a negative direction, it may be concluded that limited reinforcement may be a poor method for generating speaking improvement, at least, gestural improvement. With the PPS (a measure of general physical behavior) changing negatively, the MNW may not only have failed to encourage gestural improvement, but may have actually generated a general deterioration of behavior.

Earlier, the MXW condition showed an increase in the mean and median of the GES over the two evaluation periods.

Table 18. Wilcoxon Z-Scores for Gestural Evaluation Score and Physical Performance Score, Maximum Written Reinforcement

	GES	PPS
T	(13-)* 198	(14+)* 237.5
N	31	31
Z	.9798	.2058
p	.1635 n.s.	.3591 n.s.

*Smaller sum of like-signed ranks

Table 18 confirms that trend for the GES and the rather random movement of the PPS evaluation. Apparently, the Ss in

the MXW were provided with more information than the Ss in the MNW about their gestural performance by the EE and were able to convert that information into improved gesturing. It is interesting to note that the variability of the MXW group decreased between T_1 and T_2, while the mean and median were increasing. The standard deviation for the GES at T_1 was 43.42 and changed to 38.84 at T_2, while the variability of the PPS decreased from 20.60 to 17.66. Apparently, the MXW condition was able to improve the performances of some of the Ss, particularly those with the lowest T_1 evaluations. Perhaps this condition, with evaluations on several elements of gesturing, may have been able to supply useful information to the Ss which aided him in improving his performance behavior.

The LOR, which earlier demonstrated the greatest amount of positive change, also failed to reach statistical significance. However, the approach to the .05 level, reported in Table 19, on the Wilcoxon suggests that the magnitude of the changes in the LOR were greater than in the other two conditions.

Table 19. Wilcoxon Z-Scores for Gestural Evaluation Score and Physical Performance Score, Light On-Set Reinforcement

	GES	PPS
T	(12-)* 142	(13-)* 138.5
N	28	27
Z	1.389	1.21
p	.0823 n.s.	.1131 n.s.

*Smaller sum of like-signed ranks

While the Sign Test (testing only direction) showed a probability of .2877 (see Table 16, page 58), the Wilcoxon (testing direction and magnitude of change) produced a probability of .0823. It may be reasonable to conclude that the LOR had selective impact upon the Ss. That is, those scores which changed in a positive direction made greater increase than those scores which changed in the unpredicted direction. This conclusion may also be inferred from the increased variability that appears when comparing the GES and PPS at T_1 and T_2. The standard deviation of the GES before the experimental treatment was 22.82, while afterward the deviation rose to 32.14. On the PPS the standard deviation increased from 14.29 to 17.65. Apparently, then, the light was able to have positive impact upon some of the Ss without having serious negative effect upon other Ss.

Summary

This chapter has reported the results of the various statistical tests applied to the collected data. The purpose, of course, was to determine whether the statistical hypothesis could be rejected and the theoretic hypothesis confirmed. Prior to the actual testing of the hypotheses, several preliminary assumptions were examined. All preliminary investigations affirmed the pre-research assumptions except for the heterogeneity of variance found at T_1 in both the GES and PPS. Therefore, nonparametric statistics were computed for change scores between conditions and for raw scores within conditions. In either case, no statistical significance was found, and the theoretic hypothesis could

not be affirmed. However, it seemed apparent that changes were taking place within the various conditions. Examination of the measures of central tendency and variability seemed to confirm movement within the groups, as did directional and change analysis. That is, the positive reinforcements supplied in the MXW and the LOR had an impact upon some of the Ss. Further research seems necessary to draw definitive conclusions concerning the use of positive reinforcement in the classroom. It would seem that light could be used as supplemental positive reinforcer for gesturing and other delivery skills by the classroom instructor.

CHAPTER IV

CONCLUSIONS AND IMPLICATIONS FOR FURTHER RESEARCH

This study has compared three modes of positive reinforcement within a beginning speech class. Research in the field has concluded that reinforcement, positive and contiguous to the behavior desired, enhances learning of that behavior. It was hypothesized in this study that the MNW, MXW and LOR would show differential effects upon the Ss when improvement scores were compared. However, since only the LOR met the above criteria, it was suspected that change in the gesturing of Ss would be greater in the LOR than in either written condition.

When the nonparametric analysis of variance was applied to the improvement score data, no statistical significance was found between the Ss of the three positive modes. Finding no significance, the experimenter faces alternative reasons for the results: either the theoretic basis for the predictions was faulty, or the research itself contained errors. The latter seems the more reasonable conclusion in this research for two reasons: (1) the impact of positive reinforcement has been too well documented in education psychology and irregularly supported in communications research, and (2) the trends in the data seem to support the original prediction. Therefore, possible research errors need to be

identified and recommendations for eliminating those errors suggested before definitive conclusions may be drawn about the original predictions.

Potential Sources of Research Error

One obvious source of error may be found in the rating panels. While the median reliabilities for each scale appear sufficient to warrant adding the panels together, variabilities are readily apparent (see Appendix G, page 92). An examination of the variability between the panels suggests that the majority of raters performed their rating tasks with some agreement. However, there were three raters who, when they appear in a panel, depress the total reliability of that panel. Since the Ss scores were based upon adding across these ratings, it is possible that an individual Ss score may be inflated or deflated by the appearance of these raters upon a panel.

While unreliability of a single rater may have affected the ratings of the Ss, the panels themselves may have had an effect. That is, the appearance of a rating panel in the classroom may have had a depressing effect upon the performance of the Ss. If so, this depressing effect might have greater influence upon Ss performance at rating T_2 than at T_1. At T_1, the Ss were in only their second class meeting and were giving their first speech. Fear and uncertainty may have been working to reduce the quality of the physical performance. The appearance of a panel of strange raters may only heighten that Ss fear and uncertainty. However,

the second time the panels appear in the classroom, the Ss were in the seventh drill section and were giving their sixth speech. If the fear and uncertainty had been diminishing (because of increased familiarity with peers, instructor and EE), the appearance of another panel of strange raters may have encouraged a temporary increase in fear with a consequent reduction in amount and quality of physical behavior. The result, of course, would be a lower rating at T_2 than the Ss development at that point merited. Perhaps even a lower rating at T_2 than at T_1 would result, particularly if gestural development had been slow during the treatment speeches.

Another source of error may have been the EE. Each was a qualified, experienced teacher working toward a doctoral degree in the area of Rhetoric and Public Address. However, it is possible that the EE may have inadvertently encouraged differences to arise between Ss and between treatment groups. The EE could encourage Ss variability by providing positive or negative nonverbal cues to the Ss receiving the reinforcement. Such cues may add approval to an already approving evaluation, thereby increasing the perceived reward. Of course, a negative display could reduce the impact of a positive evaluation. Whichever direction, the impact of the EE would be selective. Some of the Ss within each condition may be changing because of their interaction with the EE.

The EE may have encouraged differences between groups if the interaction was systematic. That is, if the EE appeared uninterested in his assignment or the Ss, an entire

class might react negatively to him. Or, if the evaluator seemed particularly interested in his tasks and the Ss, the class might positively respond to him with a resultant increase in their gestural development. It is possible, although only slightly, that the negative reaction of the MNW may have been due to systematic disapproval shown the classes by the EE. Post-research conversations do not suggest such a conclusion. But these conversations may not have been accurate checks. What may have seemed unimportant (and thus, unreported to the experimenter) to the instructor and the EE after the research may have been given great significance by the class members. The result of such a circumstance might be a dislike for the EE and an ignoring of his efforts to improve their gesturing.

A fourth area for potential error in this research, with a resulting failure to show significance, may have arisen from an interaction between gesture and the reinforcement mode. One continuing complaint of the instructors, EE and raters was that several of the Ss simply did not move during their performance. The reinforcement modes used in this research required that the Ss present physical movement in order to receive the reinforcement. For example, in the MNW and MXW conditions, if the Ss did not move at all, he received "No basis for comment" as the EE evaluation. In the LOR, no gesturing would result in no light flashes. In essence, if the Ss failed to move, he selectively placed himself in a "no reinforcement" condition. Such a situation would not provide an appropriate test of the effectiveness of

positive reinforcement.

Finally, another potential for error is a consequence of using the classroom --- peer feedback. The approval or disapproval shown a speaker by his peers may have greater impact upon the physical behavior than the positive reinforcement administered by the EE. Should the peers act bored or uninterested in the speech and speaker, the result may be a deterioration in the behavior of the speaker. Such disapproving feedback might (1) reduce the amount of positive reinforcement received during the treatments (because disapproving peer feedback acted as a depressant upon delivery skills), and (2) reduce the raters' evaluations (because the Ss curtailed their movements in response to the disapproval of the peers). Such a depressing effect upon physical delivery in general, and upon gesturing in particular, may have occurred more during T_2 evaluations than in previous speeches. The sixth speech was considered the most difficult by the instructors because it required the Ss to read and formulate opinions upon current affairs. While the intent of the assignment was to ensure audience interest in the topic (topics had to be chosen from the university newspaper), it is possible that the opposite effect resulted. If so, audience disinterest may have depressed gestural behavior, lowering judges' evaluations at T_2 and resulting in the inability of the improvement score data to reach statistical significance.

Of course, the preceding remarks are only speculations. There are no data to affirm or deny the assertions. However, all or any one of the factors (if actually operating)

could discourage the gesturing of the Ss, or encourage the inaccuracy of the raters' observations. In either case, greater control of these extraneous, potentially operating variables needs to be exerted in order to diminish the effect upon the judging panel evaluations.

Further Research

Given that the theoretic base of this study seems sound, but the research control of error-producing variables weak, then a next research step would be a replication with greater control. Such a study might well remove the research from the classroom and place it within an experimental framework. Ss would be selected from the general population of students and not taken as part of intact groups. Each Ss would be randomly assigned to one of four reinforcement conditions: the three described in this study and a group in which no reinforcement was given for gesturing. Comments would be made to the students concerning other aspects of delivery and content for every speech. The pre- and post-test design used in this study would divide the treatment speeches from the rated speeches.

Each speech by each speaker would be videotaped, then coded and stored. The people witnessing the speeches would be the content evaluator, the EE administering the reinforcement and the production crew necessary to record the speech. After the completion of a series of six speeches similar in assignment to those of this study, panels of raters would view the tapes and evaluate the performances.

The raters would be selected carefully and subjected to several training sessions. Furthermore, the panels would see the tapes of speakers from the four conditions randomly mixed. In other words, a panel of judges might rate the taped performances of 10 speakers in a single session: two from the MNW, two from the MXW and three each from the LOR and the no reinforcement condition. Also, some panels would see speakers in the pre-test, post-test order, while other panels would see speakers in a reversed order. No two panels would evaluate the same collection of speakers, nor would the panels know which speaking time (pre- or post-) they were evaluating.

As in the present research, difference score data using an analysis of variance statistical design seems appropriate. If overall differences should appear between the four groups, then individual cell differences could be measured by t-tests. Intracellular improvement might be measured by correlated t-tests. Such a research project should help to encourage differences to appear between the groups if, in fact, the varying modes of positive reinforcement do produce differential effects upon the gesturing that is done by student speakers.

Aside from the proposed study described above, additional research might be undertaken to examine the theoretic hypothesis: (1) Other mechanical reinforcers might be used on college students. For instance, a money-drop might be installed in a classroom lectern to drop coins for each gesture found acceptable to the experimental evaluator. It

is possible that some collegiate speakers in the current research may have found the flash of light unrewarding.
(2) Some other speaking variables also might be studied under varying reinforcement conditions. For example, verbal variables, like total output and number of nonfluencies, which have been studied already in other research, could be examined. Pitch, rate and volume could be variables possible for study under conditions similar to those described in this research.
(3) To positively reinforce the student by removing some noxious stimulus might be a possible method. For example, a disruptive noise might be presented to the Ss until he gestured correctly. Such a disruptive noise could be a continuous bell or buzzer. The use of noxious stimuli might be particularly effective for subjects who are reluctant to move the hands and arms while speaking. The presentation of a disruptive stimulus might provide sufficient motivation for the student to move to reduce the irritation. Of course, this kind of reinforcement might not be acceptable in the classroom.

Conclusions

While no statistically signigicant results were produced by the reinforcement conditions of this study, several tentative conclusions may be drawn which supply a basis for further research: (1) positive reinforcement is recommended for speech classroom evaluation, but does not appear to be widely used; (2) immediate reinforcement is recommended to maximize skill development, but does not appear to be widely used in the speech classroom; (3) the gestures of those

students exposed to a mode of reinforcement (light on-set) meeting these recommendations tended to show greater improvement than the gestures of students exposed to other positive reinforcement modes; (4) a single evaluation of gesturing appears to be an insufficient mode for administering positive reinforcement, and (5) additional research within an experimental setting is needed to examine the impact of positive reinforcement as a teaching tool or a supplemental critique technique for the speech instructor.

Summary

The four potential studies previously mentioned and the tentative conclusions drawn indicate additional study is needed of the theoretic hypothesis of this exploratory research. With the continuous growth of higher education and the increasing enrollment size of many basic courses, there appears to be real need to implement new ways for encouraging the acquisition and refinement of students' knowledge and skills. Perhaps the variables studied herein may eventually assist speech instructors in developing the abilities of the individual students they teach.

BIBLIOGRAPHY

BIBLIOGRAPHY

Amato, Philip P., and Terry H. Ostermeier. "The Effect of Audience Feedback on the Beginning Public Speaker," The Speech Teacher, XVI (January, 1967), 56-60.

Becker, Sam. "Research on Speech Pedagogy" from Dimensions of Rhetorical Scholarship, Roger E. Nebergall (ed.). publication of the Department of Speech, University of Oklahoma, 1963, 31-48.

Berelson, Bernard, and Gary Steiner. Human Behavior: An Inventory of Scientific Finds. New York: Harcourt, Brace, and World, Inc., 1964.

Berlo, David C. The Process of Communication. New York: Holt, Rinehart and Winston, 1960.

Bigge, Morris L. Learning Theories for Teachers. New York: Harper and Row, Publishers, 1964.

Blubaugh, Jon A. "The Effects of Positive and Negative Audience Feedback on Selected Variables of Speech Behavior of Normal-Speaking College Students." Unpublished Ph.D. dissertation, University of Kansas, 1966, abstracted by C. E. Larson for Speech Monographs, XXXIV (August, 1967), p. 243.

Bostrom, Robert N. "Classroom Criticism and Speech Attitudes," Central States Speech Journal, XIV (February, 1963), 27-32.

Byers, Burton H. "Speech and the Principles of Learning," The Speech Teacher, XII (March, 1963), 136-140.

Cahn, Dudley. "The Effect of Simulated Audience Response on Speaker Attitudes." Unpublished Master's thesis, Northwestern University, 1966.

Cerekwicki, Louise E., and David Grant. "Delay of Positive Reinforcement in Instrument Eyelid Conditioning," Journal of Experimental Psychology, LXXV (1967), 360-364.

Combs, Walter, and Gerald Miller. "The Effect of Audience Feedback on the Beginning Public Speaker. I. A Counter-view," The Speech Teacher, XVII (September, 1968), 229-231.

Davis, James W. "Variations in Verbal Behavior in Dyads as a Function of Varied Reinforcing Conditions," Speech Monographs, XXXIV (November, 1967), 443-447.

Dedmon, Donald N. "Criticizing Student Speeches: Philosophy and Principles," Central States Speech Journal, XVIII (November, 1967), 276-284.

Densmore, G. E. "The Teaching of Speech Delivery," Quarterly Journal of Speech, XXXII (February, 1946), 67-71.

Deutschmann, P. J. "The Sign-Situation Classification of Human Communication," The Journal of Communication. (Mimeographed).

Eisenson, Jon, J. Jeffery Auer, and John V. Irwin. The Psychology of Speech. New York: Appleton-Century-Crofts, 1963.

Ferguson, George A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill Book Company, Inc., 1959.

Greenspoon, Joel. "The Effect of Two Nonverbal Stimuli on the Frequency of Members of Two Verbal Classes," American Psychologist, IX (August, 1954), p. 384.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill Book Company, Inc., 1954.

Harms, L. S. "Programmed Learning for the Field of Speech," The Speech Teacher, X (September, 1961), 215-219.

Hayworth, Donald (ed.). A Research into the Teaching of Public Speaking. Official Project #7255, Works Progress Administration, published by the Committee on Cooperative Research of the National Association of Teachers of Speech, Michigan Department of Public Instruction and Michigan Society of Teachers of Speech, 1939.

Hilgard, Ernest R., and Gordon H. Bower. Theories of Learning. New York: Appleton-Century-Crofts, 3rd. ed., 1966.

Hinde, Dorothy Mae. "An Investigation into the Effectiveness of Criticism of Speech Delivery When Given Before, During, and After Speech Performance." Unpublished Master's thesis, State University of Iowa, 1949.

Huenergardt, Douglas W. "An Experimental Study of the Effects of Increasing Percentages of Simultaneous Noncontingent Audience Approval on Speaker Attitudes." Unpublished Ph.D. dissertation, Northwestern University, 1967.

Johnson, F. Craig, and George Klare. "Feedback: Principles and Analogies," Journal of Communication, XII (1962), 150-159.

Karns, Charles F. "The Verbal Behavior of a Speaker as a Function of Certain Non-Verbal Aversive Stimuli Presented by an Audience in a Public Speaking Situation," Unpublished Ph.D. dissertation, University of Pittsburgh, 1964.

Kerlinger, Fred N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1965.

Kimble, Gregory A. H. (rev. ed.). Conditioning and Learning. New York: Appleton-Century-Crofts, 2nd. ed., 1961.

Krasner, Leonard. "Studies of the Conditioning of Verbal Behavior," Psychological Bulletin, LV (1958), 148-170.

Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Company, 1956.

Lockard, Robert B. "Effects of Light Upon the Behavior of Rodents," Psychological Bulletin, LX (November, 1963), 509-529.

Lothers, William Theron. "A Survey Study of Methods and Rationales in the Criticism of Classroom Speeches in the Beginning College Course." Unpublished Ph.D. dissertation, Michigan State University, 1966.

McNemar, Quinn. Psychological Statistics. New York: John Wiley and Sons, Inc., 3rd. ed., 1962.

Miller, Gerald R. "Variations in the Verbal Behavior of a Second Speaker as a Function of Varying Audience Responses," Speech Monographs, XXXI (June, 1964), 109-115.

_____, John W. Vlandis, Harry Zavos, and Milton E. Rosenbaum. "The Effect of Differential Reward on Speech Patterns," Speech Monographs, XXVIII (March, 1961), 9-16.

Ostermeier, Terry H., and Philip P. Amato. "The Effect of Audience Feedback on the Beginning Public Speaker. II. A Rejoinder," The Speech Teacher, XVII (September, 1968), 231-234.

Postman, Leo. "The History and Present Status of the Law of Effect," Psychological Bulletin, XLIV (November, 1947), 489-563.

Resnick, Lauren B. "Programmed Instruction and the Teaching of Complex Intellectual Skills: Problems and Prospects," Harvard Educational Review, XXXIII (Fall, 1963), 439-471.

Robinson, Karl F., and E. J. Kerikas. Teaching Speech: Methods and Materials. New York: David McKay Co., Inc., 1963.

Sereno, Kenneth K. "Changes in Verbal Behavior of Females During Two Successive Speech Performance as a Function of the Sequence of Listener Responses." Unpublished Ph.D. dissertation, University of Washington, 1964.

Shelton, Ralph L., William B. Arndt, and June B. Miller. "Learning Principles and Teaching of Speech Language," Journal of Speech and Hearing Disorders, XXVI (November, 1961), 368-376.

Sidman, Murray. "Operant Techniques," Chapter 6 of Experimental Foundations of Clinical Psychology. Arthur J. Bachrach, (ed.). New York: Basic Books, Inc., 1962, 171-210.

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, Inc., 1956.

Skinner, B. F. Science and Human Behavior. New York: The Free Press, 1953.

Smith, Donald K. "What Are the Contemporary Trends in Teaching Speech?" The Speech Teacher, X (March, 1961), 87-94.

Smith, Raymond G. "The Criticism of Speeches: A Dialectical Approach," The Speech Teacher, X (January, 1961), 59-62.

Vlandis, John W. "Variation in the Verbal Behavior of a Second Speaker as a Function of Varying Audience Responses," Speech Monographs, XXXI (June, 1964), 116-119.

Wallace, Murray R. "Techniques and Devices in Teaching Speech," The Bulletin of the National Association of Secondary-School Principals, XXXVIII (January, 1954), 210-214.

Weaver, Andrew T., Gladys L. Borchers, and Donald K. Smith. The Teaching of Speech. New York: Prentice-Hall, Inc., 1952.

Wiener, Norbert. The Human Use of Human Beings. New York: Doubleday Anchor Books, 1954.

Winer, B. J. Statistical Principles of Experimental Design. New York: McGraw-Hill Book Company, 1962.

APPENDICES

APPENDIX A

Questionnaires to Faculty and Teaching Assistants

Dear Colleague:

Although you are often asked to assist in research, I hope you will give me a few moments of your time.

We are interested in the gestural behavior of beginning public address students. In your evaluations of beginning speakers, what qualities of gestural behavior do you look for? To ask the same question in a different manner, what qualities of gesturing do you wish to develop in your beginning speech students?

Please indicate as many of these qualities as you think important in the consideration of the beginning speech student. And, why not do it soon? Remove this little task to be free from tensions during the holidays.

Thank you for your cooperation. Have a delightful Christmas Holiday. You may place these completed forms in my mailbox in the Speech Office.

Dan P. Millar

Dear Colleague:

Once again I ask you to assist me in my research. I know these requests take time, but I hope you will spend a few moments and answer the following questions.

The following is a set of words and phrases used to describe good qualities of speaker gesturing (the hand and arm movements of the speaker). Please look carefully at the list and determine if each represents a separate and distinct category of gestural evaluation.

1. Do you perceive the qualities listed below as separate and distinct categories? (please circle)

   Yes No

2. If you answered "No," please look at the words and phrases again. If you see similarities or overlappings, group the overlapping (similar) words and phrases together in the space below.

LIST OF QUALITIES

visibility	appropriate to content
natural	coordinated with other bodily action
fits the speaker	free
clarity	definite
forceful	complete
emphatic	directed toward the audience
spontaneous	relaxed

Are there any qualities of gesturing you find important and missing from this list? If so, please indicate:

Thank you very much for your continued cooperation. Please place the completed form in my mailbox.

Dan P. Millar

APPENDIX B

Gestural Evaluation Instrument and Rater Instructions

INSTRUCTIONS TO RATERS

Dear _____:

Attached is a sample of the instrument to record your evaluations of the physical behavior of student speakers. The upper portion of the sheet refers only to the gesturing of the student (hand and arm movements). The lower portion, beginning with "General Impressions. . . ," asks you to evaluate the total physical performance (walking, posture, head and trunk movements, etc.) of the student speaker.

Remember, these are students in the beginning speech course and should be evaluated accordingly. Further, you should not relay your evaluations to other judges, the instructor or the students. This sheet should be an independent measure of your evaluation.

The class instructor will call and spell the name of the speaking student. Please write the last name in the upper right-hand corner. Wait for the speech to be concluded before you mark your evaluation. Keep your evaluations together and return them to me in 110 South Hall when the class period has ended. If the speeches are completed before the period ends, please wait until the instructor asks you to leave before you go.

Please meet in 110 South Hall approximately 10 minutes before the start of the class period. Any questions you may have can be answered at that time, or earlier, if necessary. You may carry these instructions with you into the class.

Thank you very much for assisting me in this research. I am sure you realize how important this work is to me.

_____(Time and Place)_____

_____(Instructor's Name)_____

GESTURE EVALUATION

Speech 102 Spring Quarter

The following are elements of gesturing. You are being asked to carefully evaluate the gestural behavior of the speakers you will see and hear. Be sure you evaluate the speaker on every element and then mark your evaluation in the appropriate blank for <u>each</u> <u>and</u> <u>every</u> <u>element</u>.

VISIBILITY

visible invisible

:_____:_____:_____:_____:_____:_____:_____:

CLARITY

distinct indistinct

:_____:_____:_____:_____:_____:_____:_____:

DEFINITE

indefinite definite

:_____:_____:_____:_____:_____:_____:_____:

COMPLETE

complete incomplete

:_____:_____:_____:_____:_____:_____:_____:

FITS THE SPEAKER

natural unnatural

:_____:_____:_____:_____:_____:_____:_____:

SPONTANEOUS

restricted free

:_____:_____:_____:_____:_____:_____:_____:

APPROPRIATE TO THE CONTENT

appropriate inappropriate

:_____:_____:_____:_____:_____:_____:_____:

COORDINATED WITH OTHER BODILY ACTION

coordinated uncoordinated

:_____:_____:_____:_____:_____:_____:_____:

DIRECTED TOWARD THE AUDIENCE

undirected directed

:_____:_____:_____:_____:_____:_____:_____:

GENERAL IMPRESSION OF PHYSICAL PERFORMANCE

bold timid

:_____:_____:_____:_____:_____:_____:_____:

aggressive meek

:_____:_____:_____:_____:_____:_____:_____:

tired energetic

:_____:_____:_____:_____:_____:_____:_____:

confident unsure

:_____:_____:_____:_____:_____:_____:_____:

colorful dull

:_____:_____:_____:_____:_____:_____:_____:

APPENDIX C

Instruments of the Minimum Written Reinforcement Condition

Speech 102

The following is an element of good gesturing. The hand and arm movements that you made during your speech have been evaluated and this element is considered the best element of your gestural performance.

VISIBILITY

Excellent-- your gestures were easily seen	Good--could see most of your gestures	Okay--could see a few of your gestures	No basis for comment

. . . . . .

Speech 102

The following is an element of good gesturing. The hand and arm movements that you made during your speech have been evaluated and this element is considered the best element of your gestural performance.

CLARITY

Excellent-- your gestures were distinct	Good--most gestures were distinct	Okay--a few gestures were distinct	No basis for comment

. . . . . .

Speech 102

The following is an element of good gesturing. The hand and arm movements that you made during your speech have been evaluated and this element is considered the best element of your gestural performance.

DEFINITE

Excellent-- your gestures were forceful and emphatic	Good--most gestures were forceful and emphatic	Okay--a few gestures were forceful and emphatic	No basis for comment

. . . . . .

Speech 102

The following is an element of good gesturing. The hand and arm movements that you made during your speech have been evaluated and this element is considered the best element of your gestural performance.

COMPLETE

Excellent-- each gesture was finished	Good--most gestures were finished	Okay--a few gestures were finished	No basis for comment

Speech 102

The following is an element of good gesturing. The hand and arm movements that you made during your speech have been evaluated and this element is considered the best element of your gestural performance.

FITS THE SPEAKER

Excellent-- each gesture was natural	Good--most gestures were natural	Okay--a few gestures were natural	No basis for comment

. . . . . .

Speech 102

The following is an element of good gesturing. The hand and arm movements that you made during your speech have been evaluated and this element is considered the best element of your gestural performance.

SPONTANEOUS

Excellent-- each gesture was relaxed and free	Good--most gestures were relaxed and free	Okay--a few gestures were relaxed and free	No basis for comment

. . . . . .

Speech 102

The following is an element of good gesturing. The hand and arm movements that you made during your speech have been evaluated and this element is considered the best element of your gestural performance.

APPROPRIATE TO THE CONTENT

Excellent-- each gesture related to the thought	Good--most gestures related to the thought	Okay--a few gestures related to the thought	No basis for comment

Speech 102

The following is an element of good gesturing. The hand and arm movements that you made during your speech have been evaluated and this element is considered the best element of your gestural performance.

COORDINATED WITH OTHER BODILY ACTION

Excellent--each gesture was coordinated with other movements	Good--most gestures were coordinated with other movements	Okay--a few gestures were coordinated with other movements	No basis for comment

. . . . . .

Speech 102

The following is an element of good gesturing. The hand and arm movements that you made during your speech have been evaluated and this element is considered the best element of your gestural performance.

DIRECTED TOWARD THE AUDIENCE

Excellent--each gesture was directed to us	Good--most gestures were directed to us	Okay--a few gestures were directed to us	No basis for comment

APPENDIX D

Instrument of the Maximum Written Reinforcement Condition

GESTURE EVALUATION

Speech 102 Spring Quarter

The following are elements of good gesturing. The hand and arm movements that you made during your speech have been evaluated on the basis of these elements and the evaluation indicated.

1. VISIBILITY			
Excellent--your gestures were easily seen	Good--could see most of your gestures	Okay--could see a few of your gestures	No basis for comment
2. CLARITY			
Excellent--your gestures were distinct	Good--most gestures were distinct	Okay--a few gestures were distinct	No basis for comment
3. DEFINITE			
Excellent--your gestures were forceful and emphatic	Good--most gestures were forceful and emphatic	Okay--a few gestures were forceful and emphatic	No basis for comment
4. COMPLETE			
Excellent--each gesture was finished	Good--most gestures were finished	Okay--a few gestures were finished	No basis for comment
5. FITS THE SPEAKER			
Excellent--each gesture was natural	Good--most gestures were natural	Okay--a few gestures were natural	No basis for comment
6. SPONTANEOUS			
Excellent--each gesture was relaxed and free	Good--most gestures were relaxed and free	Okay--a few gestures were relaxed and free	No basis for comment
7. APPROPRIATE TO THE CONTENT			
Excellent--each gesture related to the thought	Good--most gestures related to the thought	Okay--a few gestures related to the thought	No basis for comment
8. COORDINATED WITH OTHER BODILY ACTION			
Excellent--each gesture was coordinated with other movements	Good--most gestures were coordinated with other movements	Okay--a few gestures were coordinated with other movements	No basis for comment
9. DIRECTED TOWARD THE AUDIENCE			
Excellent--each gesture was directed to us	Good--most gestures were directed to us	Okay--a few gestures were directed to us	No basis for comment

APPENDIX E

Speaking Assignments

ASSIGNMENTS

Speech 102 Spring Quarter

During the first few class periods of this quarter, you will give six speeches. This will be different from the other sections of Speech 102. However, the differences will not affect you (or your grade) and are simply part of the continuing changes made necessary by growing enrollments and television teaching.

The following are the assignments for your drill section periods only:

Drill Section Period	Assignment
first	Introduction to the course.
second	A speech of self-introduction. Focus upon things you like to do, future plans, things you have done, etc. Do NOT talk just about your high school. Example: your major and your plans for jobs; a hobby and why you like that specific activity.
third	A speech of information based upon your reading of current news magazines. Read a major article from one of the following news magazines: National Observer, Newsweek, Saturday Review, U.S. News & World Report and Time. Look only at issues from within the past 3 weeks. The object is to take this article and report upon the topic in class.
fourth	A speech of problem-solution on some local (campus or city) problem. Take a local problem, think about it and attempt to formulate a constructive program for solving it. This is not a research assignment but a thinking one. Example: campus parking, lack of married housing, student government.
fifth	A speech of information based upon your vocational field. Look at your vocational field and discover the influences and trends and directions. Read from scholarly journals, talk to your professors to discover

what trends exist in your field and how these trends will influence your vocation in the future. Example: the current learning research and what effect this knowledge will have upon the teaching of retarded (or accelerated) students in the next 10-20 years; or, the impact of computer technology upon business.

sixth — A speech of entertainment based upon your summer experiences. Using your job, travels, dates, play-time as a base for this speech, build an oral presentation that the class will find enjoyable and amusing.

seventh — A speech of information or persuasion based upon current events as reported in the BG News. Read an article from the BG News, expand upon the topic, develop it for the class and present it. You may use national, state, city events or problems, campus events and problems including play reviews, sporting news, movie reports, etc. But this is a research assignment so you will want to go beyond the BG News to find additional views and facts. Notice that you may simply report to the class on your topic or you may try to convince them to take a new stand (adopt a new opinion) on the topic. Example: the success of the baseball team. Talk with players, coaches, etc., and then deliver a persuasive speech advocating the campus support the team by attending the games.

VERY IMPORTANT--READ. The following regulations are placed upon the speeches you will give during the first seven drill section periods. And those sections only. Other assignments will be made for later drill section periods. Please observe these regulations.

1. Minimum time of 2½ minutes and a maximum of 4 minutes. If you speak less, evaluation of your performance is difficult and inaccurate. If you speak longer, some one else will be limited in his time because of the

number of students in the class.

2. Only those speeches meeting the above assignments will be acceptable.

3. All speeches must be original. A great part of public speaking is learning to create topics and organize ideas.

4. You may use notes while you speak. Practice using them so you can relax and speak directly to the audience.

5. No written assignment is required for the speeches given during the first seven drill section periods.

Each speech you give will be evaluated and critical comments made concerning your performance.

APPENDIX F

Instructions to Instructors

Carol, Donna, Greg:

(MNW)

Please begin your class on Tues. this way: "The gentleman visiting our class is a member of the teaching faculty of the Department of Speech. He is here to use a new form for gestural evaluation. I will evaluate your speech in its entirety; he will evaluate only gestures. Please go to him at the end of your speech to receive his written evaluation."

(MXW)

Please begin your Wed. class this way: "The gentleman visiting our class is a member of the teaching faculty of the Department of Speech. He is here to use a new form for gestural evaluation. I will evaluate your speech in its entirety; he will evaluate only gestures. Please go to him at the end of your speech to receive his written evaluation."

(LOR)

Please begin your Wed. noon class this way: "The gentleman visiting our class is a member of the teaching faculty of the Department of Speech. He is here to use a new method for gestural evaluation. The box in front contains a light. Whenever you see that light flash, it means that the gesture you have just made is considered good --- the gesture that you just made was considered <u>good</u>. I will evaluate your speech in its entirety."

Thanks. Questions?

MEMORANDUM May 6, 1969

TO: Carol Shaw, Donna Emlich, Greg Gardner
FROM: Millar

Thank you for permitting me to use your classes for my dissertation experiment. I know this was an inconvenience and may yet cause you considerable discomfort with your classes. I hope any damage done will not be irreparable, however, and that your classes may assume some semblance of normalcy before the end of the quarter.

One last favor. Please see me at your convenience this week. I would like to talk with each of you for a short time about (1) what you observed in the classroom, (2) any student feedback, and (3) your own feelings concerning what you saw and the intent of the experiment.

The following you may read to your classes if you wish. They may be somewhat relieved to hear what was going on:

"To the 102 classes involved in the experiment:

Thank you for giving your time and energies to help 102 continue to grow and develop. You were subjected to written and mechanical devices designed to help you better understand your arm and hand movements while speaking. The objectives of the experimentation were the following:

(1) to see if a many-speaking experience, without prior training, was a superior approach to 102 than other approaches.
(2) to see if teachers of 102 could reasonably agree on the characteristics of good gesturing.
(3) to see if 102 could use two evaluators in order to improve the performance of speakers: one to look at some specific area, in this case gesture, but freed from the subjective stigma of giving a grade, and the other to look and grade.
(4) to see if written forms dealing with some specific aspect of speaking are superior to the rather general comments made by an instructor.
(5) to see if some mechanical means of criticism, in this case the light, would encourage the generation of some specific desirable speaking trait, in this case gesturing.
(6) to see if 'strangers' to the classroom influence negatively or positively the behavior of beginning students. From a speaking concern, it might be desirable to have a single class evaluated by a variety of instructors during the course of a quarter.

If you, as students, have comments of either a negative or a positive nature concerning your experience during the first few class meetings, please pass them along to your instructor. Or, if you wish to remain anonymous, write them to Professor Delmer Hilyard, Co-Director of 102, Department of Speech."

Thanks.

APPENDIX G

Table 20. Rater Reliability for Gestural Evaluation Score, Time 1 and 2

and

Table 21. Rater Reliability for Physical Performance Score, Time 1 and 2

Table 20. Rater Reliability for Gestural Evaluation Score, Time 1 and 2

VISIBLE		CLARITY	
Time 1	Time 2	Time 1	Time 2
1.0000	.9137	.9259	.9005
.9014	.8660	.8430	.7719
.8778	.7976	.7760	.7466
.8180	.7167	.7714	.7148
.8012	.7005	.7424	.7136
.7241	.6935	.7222	.6365
.4329	.4405	.6313	.5570
.4006	.3613	.2846	.5123
.2522	.0000	.0004	.0456

DEFINITE		COMPLETE	
Time 1	Time 2	Time 1	Time 2
.9326	.9482	.9773	.8459
.8542	.9248	.9739	.8200
.7545	.8851	.8551	.8088
.7187	.6790	.8324	.7659
.6676	.6481	.5856	.7533
.6134	.5222	.4837	.6470
.1685	.4459	.4832	.6382
.1117	.4455	.3363	.6382
.0034	.0053	.1630	.6307

NATURAL		FREE	
Time 1	Time 2	Time 1	Time 2
1.0000	.9050	.9961	1.0000
.8670	.8572	.8930	.9629
.7971	.6722	.8525	.8923
.7878	.6434	.6745	.8847
.7595	.5566	.6598	.8265
.7592	.4923	.6235	.7709
.6204	.3784	.5837	.7359
.5898	.2755	.0005	.4347
.3823	.2561	.0000	.3874

Table 20. (cont'd.)

APPROPRIATE TO THE CONTENT		COORDINATED WITH BODILY ACTION	
Time 1	Time 2	Time 1	Time 2
.9966	.8152	1.0000	.9329
.9096	.8032	.9813	.8751
.8100	.8006	.7539	.8562
.6370	.6364	.7473	.7544
.6058	.5760	.6390	.6715
.5329	.5499	.6333	.6689
.5152	.4331	.6148	.6378
.4355	.3070	.4273	.5881
.4271	.2311	.3867	.2555

DIRECTED TO THE AUDIENCE		TOTAL GES	
Time 1	Time 2	Time 1	Time 2
.9362	.8338	.9987	.9381
.9033	.7682	.9880	.9127
.8701	.7401	.9650	.8032
.6761	.7224	.8790	.7321
.6633	.7142	.8355	.6777
.5788	.6059	.7075	.6377
.5525	.5411	.6405	.5426
.5465	.5329	.6041	.3316
.4467	.4300	.5544	.2041

Table 21. Rater Reliability for Physical Performance Score, Time 1 and 2

BOLD		AGGRESSIVE	
Time 1	Time 2	Time 1	Time 2
.9022	.8500	.8993	.8775
.8674	.8047	.8142	.8596
.8541	.7985	.7853	.8132
.8168	.7631	.6923	.7820
.7702	.7469	.6391	.7632
.6083	.7334	.6315	.6616
.2463	.7280	.4002	.6544
.1117	.6931	.2558	.5602
.0750	.6344	.0840	.5459

ENERGETIC		CONFIDENT	
Time 1	Time 2	Time 1	Time 2
.8993	.8670	.9032	.8801
.8762	.8503	.8400	.8223
.7403	.8437	.7864	.7560
.6795	.8304	.7286	.7499
.6128	.6728	.6164	.7421
.6078	.7269	.5504	.7383
.4435	.5548	.4764	.7266
.3773	.5433	.1386	.7211
.1122	.4587	.0910	.4851

COLORFUL		TOTAL PPS	
Time 1	Time 2	Time 1	Time 2
.9533	.8686	.9388	.9125
.8379	.8428	.8771	.8526
.8288	.8269	.8592	.8433
.7987	.8184	.7495	.8249
.7663	.8140	.5175	.7790
.7660	.7288	.4069	.6984
.6900	.7020	.3889	.5879
.5058	.6962	.3313	.4202
.3029	.0891	.2596	.3802