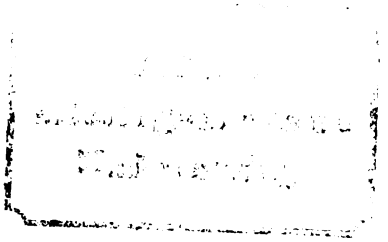
THE BIASING EFFECT OF VISUAL CUES IN LANGUAGE EVALUATIONS

Thesis for the Degree of M. A.
MICHIGAN STATE UNIVERSITY
KATHRYN ZIMMERMAN HARLTON
1972



ABSTRACT

THE BIASING EFFECT OF VISUAL CUES

IN LANGUAGE EVALUATIONS

By

Kathryn Zimmerman Harlton

The phenomenon of experimenter bias has been recognized and studied by investigators in both the physical and social sciences. Experimenter bias has been defined as the influence pre-information has upon the experimenter's evaluation of a subject's present performance. This bias has been shown to exist in the fields of Education, Sociology and Psychology, in both laboratory and non-laboratory environments. However, there has been little investigation of the phenomenon of experimenter bias in the areas of speech pathology and audiology. The purpose of this study was to investigate the possible existence of experimenter bias in the evaluation of children's language samples, utilizing visual and auditory pre-information.

The evaluators used in this study were fifteen audiologists and fifteen speech pathologists, all of whom were master's level graduate students. The evaluators were required to make subjective evaluations of a ten-minute pre-recorded language sample of a seven year old normal speaking male child. The subjective measures consisted of

two types of rating scales. The first type required rating four language attributes (structural sophistication, grammatical accuracy, creativity, and content) on a seven-point continuum. The second type used five descriptive attributes (like-dislike, mature-immature, good-bad, pleasant-unpleasant, and intelligible-unintelligible) on a seven-point continuum.

The thirty evaluators were assigned to three ten-member experimental conditions, (five audiologists and five speech pathologists). These three conditions differed in the type of pre-information and the attached photograph (visual cue) which the evaluator was given prior to listening to the speaker. The types of pre-information were termed: negative information (NI), positive information (PI), and lack of information (LI). The pre-information consisted of a folder containing contrived descriptions and information on behavior and intellectual achievement similar to the type that is found in clinical files. The photograph of the child was attached to the folder. All of the evaluators performed their tasks individually under controlled experimental conditions.

The results of the study revealed no strong biasing effects among the audiologists and speech pathologists. However, it was found that the audiologists tended to be influenced more than speech pathologists, and that this influence was in the predicted direction for both scales. The audiologists tended to rate the speaker better under the positive information condition. The speech pathologists, on the other hand, tended to rate the speaker better under the negative information condition.

If the situation were reversed however, and the speech pathologists were engaged in an audiological task, it can be hypothesized that the speech pathologists would tend to perform the task in a predicted direction.

At this point in the investigation of experimenter bias with regard to speech pathology and audiology it can be speculated that case history information does not have an effect on the evaluator. However, suggestions are presented for future research studies which involve direct interaction between the evaluator and the child in order to discover possible subtle biases occurring in speech pathology and audiology.

THE BIASING EFFECT OF VISUAL CUES IN LANGUAGE EVALUATIONS

By

Kathryn Zimmerman Harlton

A THESIS

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

MASTER OF ARTS

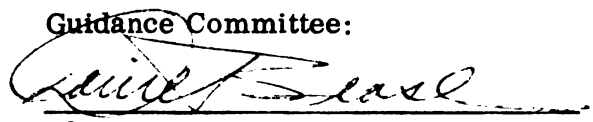
Department of Audiology and Speech Sciences

1972

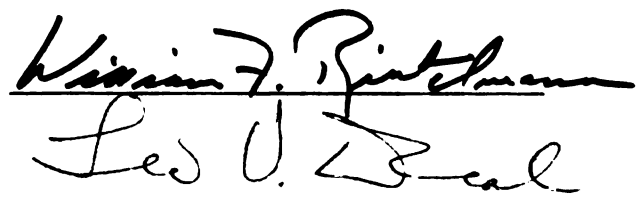
Accepted by the faculty of the Department of Audiology and Speech Sciences, College of Communication Arts, Michigan State University, in partial fulfillment of the requirements for the Master of Arts degree.

Director of Thesis

Guidance Committee:



Chairman



ACKNOWLEDGMENTS

I would like to thank Dr. Daniel S. Beasley, my thesis director, and Dr. William F. Rintelmann and Dr. Leo V. Deal, my committee members, for their kind assistance and helpful suggestions.

In addition, I would like to extend my grateful appreciation to my husband, Bruce and my parents Mr. and Mrs. John L. Zimmerman for their kind encouragement, thoughtfulness, and support. The interest and support of Mr. and Mrs. William F. Harlton, Jr. is also much appreciated.

Finally, I would like to thank Kay Kiomento for all her time and effort, and a special thanks to Pat Larkins for her encouragement, thoughtfulness, and friendship.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Rosenthal (1966) has defined experimenter bias as the extent to which experimenter effect or error is asymmetrically distributed about the "correct" or "true" value. It is a measure of net error, and generally refers to the preconceived ideas the experimenter has about his subject and his subject's performance. For the past decade investigators in the physical and social sciences have studied various aspects of the concept of experimenter bias. Johnson (1953) states that "Our assumptions define and limit what we see, i.e., we tend to see things in such a way that they will fit in with our assumptions, even if this involves omissions or distortions." (p. 79)

Experimenter Bias

When studying experimenter effect and experimenter bias, it is necessary to consider several distinguishing concepts. Rosenthal (1966) defined experimenter effect as the extent to which the data obtained by an experimenter deviate from the "correct" value. The measure of experimenter effect or experimenter error is some function of the sum of the absolute deviations of that experimenter's data about the "correct" value. Thus, experimenter effect is a refection of "the degree to which the tester influences a person's test results." (Hipskind and Rintelmann, 1969, p. 298).

This influence can be subdivided into active and passive effects. Active effects are those associated with unintended differences in the experimenter's behavior that can influence the behavior of the subject. Passive effects, on the other hand, are associated with the experimenter's appearance rather than his behavior. Rosenthal (1966) stated that distinguishing between active and passive effects is very difficult, and there have been no experiments reported to aid in making this distinction.

In addition to the above can be added the consideration of the "self-fulfilling prophesies" and "demand characteristics" associated with experimenter effects. The concept of the "self-fulfilling prophecy" was developed by Merton (1948) and reflects upon a person's expectation of an event. This expectation changes the behavior of the person in such a way as to make the prophesied event more likely to occur. The reality of the self-fulfilling prophecy has been shown in education in the form of the effect of teacher expectancy of a child's intellectual performance on that child's actual performance. MacKinnon (1962), as reported by Rosenthal (1966), feels that if it is expected that a child of a certain intelligence will not respond positively to a given task and if this expectation is known to him, the probability that he will respond positively is greatly reduced.

"Demand characteristics" refer to the evaluator's expectations or biases as conveyed to the subject by certain cues which shape the subject's

behavior (Orne, 1962). Orne has demonstrated that a variety of experiments are performed by subjects in a manner that they believe they should perform. If a subject believes that catalepsy of the dominant hand will result when hypnotized, then such a reaction will occur when hypnosis is performed. If they are not led to believe that catalepsy is a part of hypnosis, they will not show this reaction when hypnotized (Orne, 1969). While self-fulfilling prophesies deal with expectations of a subject's behavior by an experimenter, demand characteristics deal with the subject's expectations of himself and the experimenter. Thus, the subject may feel obligated to be a good subject and please the experimenter. Orne (1962) suggested that the subject may feel he must "validate the experimental hypothesis" as revealed to him previously by the experimenter. Beez (1968) suggested the possibility that "demand characteristics" may unintentionally reveal to the subject just what the experimenter is investigating and what he expects from his subject.

One area in which experimenter bias has been studied is the field of animal psychology. Rosenthal and Fode (1963) performed a study in which they labeled an experimenter's rat as "bright" or "dull." The actual performance of the rat when tunneling a T-maze was affected when the rat was tagged with a particular label. The experimenters who believed their rats were "bright" found that their rats performed better and improved daily. The other group of experimenters who believed their rats to be "dull" did not see the same results. Cardaro

and Ison (1963) studied the turning behavior of planaria. They asked the experimenters to watch for the turning. The experimenters who were told to expect more turning by their planaria recorded more turning. Inadvertent recording errors by the observer (Rosenthal, 1970) could be one possible explanation for these observations. Another explanation might be unintentional conditioning of the animals by the experimenter.

Several investigators have studied the presence of experimenter bias, in human behavior research, including how such bias would affect the results of the study, and how this bias could influence the reactions of the subjects. Rosenthal and Fode (1963) required subjects to rate photographs of a person's face on a ten point scale of success to failure. The experimenters were told that they would either obtain high "success" ratings or low "failure" ratings from their subjects. The authors reported that experimenters expecting success ratings from their subjects obtained higher mean ratings than evaluators who expected failure ratings.

Associated with the bias presented by the experimenter's preconceptions of the results he will obtain are the factors of age, sex, anxiety level and need for approval, of both the experimenter and the subject. The closer the experimenter and the subject are in age, the more influence the experimenter has over his subject's results (Rosenthal, 1966). Further, female subjects rated photographs in a person-perception task as less successful persons when the experimenter was a female rather

than a male (Rosenthal, 1966). Experimenters lower in need for approval obtained ratings of the photos as successful people. In another study involving person-perception task it was found that persons with high anxiety obtained higher ratings of success of the photographs (Rosenthal, 1966), although these results have been questioned by other investigations (Rosenthal, Kohn, Greenfield, and Carota, 1965). The less anxious experimenters obtained higher ratings of the success of the photographs.

Experimenter bias may be conveyed through both visual and auditory cues. Rosenthal (1970) reports a study by Adair and Epstein (1968) involving the person-perception task. The study was conducted in two stages. In the first part of the study the subjects both heard and saw the evaluator. In the second stage, the subjects simply heard a tape recording of the experimenter's instructions. It was anticipated by the authors that only the visual session would reveal any form of bias. However, even when the subject simply heard the tape recorded instructions, an experimenter effect was evident. Also, there was a greater biasing effect for the nonvisual condition when the experimenter expected lower ratings. In another study of the person-perception task (Rosenthal and Fode, 1963) three conditions were utilized: visual and auditory, visual only, and auditory only. For the first condition the subjects both heard and saw the experimenter. Four groups involving experimenter-subject interaction were formed in the second condition. Group I subjects saw and heard the experimenter. The experimenter had been led to expect

low ratings for all groups. Group II subjects were separated from the experimenter by a screen. A nonverbal condition was employed in group three: the subjects read the instructions while the experimenter remained silent. In Group IV the experimenters were led to expect high ratings. This group used both visual and auditory cues. The results indicated that the nonvisual group ratings came between the ratings for Group I and Group IV. The nonverbal group obtained ratings almost the same as for Group I. The authors concluded that with visual cues removed the bias was diminished. The loss of verbal cues obtained an effect similar to the non-visual group in the study conducted by Adair and Epstein.

Everyday situations have also prompted researchers to delve into the effects of evaluator bias and study how these biases are related to these life situations. For example, evaluator bias was investigated in the classroom. The teacher was considered the experimenter and the students were the subjects. A study was conducted to test the effect of teacher's self-fulfilling prophesies and to study how these prophesies affected the behavior of children from a low socioeconomic environment. (Rosenthal and Jacobson, 1968). At the beginning of the school year, the children were given The Flanagon Test of General Abilities, which consists of measures of verbal and reasoning skills. The teachers were told that this was a test of "intellectual blooming" and that some students could be expected to make high gains. Eight months later the same test was re-administered. The results showed the children who

were expected to gain in intelligence did so, and they gained more than a control group of children. The control group of children were not expected to gain in intelligence. The most affected area seemed to be the lower grades. Evans and Rosenthal (1969) replicated this study, only this time they used middle-class children. The results in this study were similar to those in the previous study. In both studies it was found that the experimenter effect was more prevalent on the reasoning tasks than on the verbal tasks. Rosenthal and Evans felt that the expectations of the teacher might have served as an unintentional self-fulfilling prophecy.

The effect of "demand characteristics" on Headstart children was studied by Beez (1968). Each experimenter was instructed to work individually with a child. They were told that the purpose of the study was to see how the child would perform on various experimental tasks. Before coming in contact with the child, the experimenter was given hypothetical data about the child. The data included background information, testing behavior, and results and interpretation of test data concerning the child. Results showed that experimenters who had favorable expectations for their child tried to teach more symbols to the child. They also rated him higher on achievement, social competence, and intellectual ability. Further, mean number of symbols learned was higher for children who were expected to achieve.

Thus, the reality of experimenter bias phenomena in psychology and

education, in general, has been demonstrated. It is presumed that such effects likely exist in the areas of speech pathology and audiology and require investigative consideration.

Experimenter Bias in Speech Pathology and Audiology

The audiology profession has been concerned with the possible biasing effects associated with previous knowledge of audiograms and scoring errors. Hipskind and Rintelmann (1969) conducted a study in which they investigated tester bias in relation to pure tone and speech reception thresholds and speech discrimination scores. Efforts were made to induce tester bias. The examiners were given one of four types of pre-information. This pre-information contained an audiogram with the subject's actual threshold, his threshold increased 10 dB, his threshold decreased 10 dB, or no information pertaining to the subject's threshold. There was no effect associated with the type of pre-information.

Nelson and Chaiklin (1970) compared two methods of scoring and scoring bias for speech discrimination testing. A write-down method involved the subject's writing down the word as he heard it. A talkback method involved the subject's repeating the word to the examiner. The examiner recorded the word as correct or incorrect. The authors felt that there might be an increased possibility of scoring bias in the talkback method. It was felt that the examiner might be questionable. Results showed that only inexperienced examiners revealed a significant difference between scores for the two scoring methods.

There have been few studies conducted in the field of speech pathology relative to experimenter bias. One of the major concerns is the evaluation of speech, hearing and language behavior. In this situation the experimenter bias may be more appropriately termed evaluator bias.

An example of a problem in speech pathology as related to evaluator bias is that of the language of the inner-city black child. The language spoken by the inner-city black child has been of recent concern to the speech pathologist. There has been controversy as to whether or not to teach "standard English" to these minority group children. There are two viewpoints characteristic of the term "nonstandard English." One view is that the child's speech is deficient relative to the norm. The child's language is termed a restricted code, thus restricting the child in cognitive skills. The other viewpoint sees the child's language as simply different from "standard English." The child's language would be considered appropriate for his particular culture and environment.

Williams (1970) felt that these "disadvantaged" children are categorized into stereotypic groups. That is, the listener stereotypes the child into a certain group from his own personal reactions to the child's speech. The language evaluator, as a listener, generalizes his reactions to the child's speech in a manner reflective of a negative "self-fulfilling prophecy." The evaluator expects a certain pattern of speech to be used by the speaker, and any language patterns that do not coincide with the prophecy can be considered incorrect by the evaluator at the outset.

Siegel stated that:

> ...ignorance of the areas of speech pathologies and language deviations may constitute an experimental safeguard against particular biases or expectations. The experimenter who is unaware that cleft palate speakers generally have a particular pattern of speech disturbances, or that certain sounds are usually mastered before others may be less likely to impose these expectations on the tests he administers. (1962, p. 34)

Unfortunately, the nature of speech pathology is such that the speech pathologist, as an evaluator of behavior, enters the testing situation with several possible biases. The fact that clients are referred to the speech pathologist immediately implies a pathology on the part of the client and is a possible source of evaluator bias (Manning and Beasley, 1971). Background information usually accompanies the referral of the client. The speech pathologist tries to become familiar with this information before seeing the client. The real danger of bias may appear when this background information includes a label such as "cleft palate," "mentally retarded," "brain damaged" (Manning and Beasley, 1971).

Siegel (1963) conducted a study dealing with interpersonal interaction of adults with retarded children. The adults were informed that the children either had high verbal skills or low verbal skills. The adults were assigned to teach a particular task to the children. The results indicated that the adults did not respond differently to the children with regard to the labels given the children. Siegel concluded that

the labels "did not significantly affect adult responses to these children" (p. 422).

A study conducted by Meitus, Ringel, House, and Hotchkiss (1972) investigated the effect of clinician bias with regard to evaluating the articulation behavior of children. Thirty clinicians viewed case presentations via videotape. Prior to the evaluation, the clinicians were given fictional case histories containing positive-bias, negative-bias, or no information. The clinicians rated phonetic inventories, judged the proficiency of articulation, and provided a prognosis and therapeutic judgment. The results indicated no significant differences in scores for all three groups. The authors feel that because the responses under the positive bias and the negative bias agree strongly, case histories are used for other reasons than determining therapy and prognosis.

Cowen, Weber, Hoddinott, and Klein (1967) conducted a study involving Mean Length of Response (MLR) as a function of stimulus, experimenter and subject. They tested ninety-six school age children from both lower and upper socioeconomic levels. The MLR measures were obtained by two of the authors. It was found that certain stimuli elicited a greater length of response. The sex, socioeconomic level, and age of the subjects interacted with the experimenter effect and the experimenter. The differences between the two experimenters' scoring and recording methods also contributed to the experimenter effect. The authors concluded that the Mean Length of Response varied with the stimulus used, the age, sex,

and socioeconomic level of the subject and the experimenter doing the evaluation.

A study involving evaluator bias and children's language performance examined objective as well as subjective measures (Manning and Beasley, 1971). The study was conducted in such a manner as to find out whether or not speech pathologists are influenced by information they receive before beginning to test a child's language abilities. The study consisted of judgments of tape recordings of two speakers. Forty evaluators were randomly assigned to one of four conditions of biasing information: negative information, positive information, incomplete information, and lack of information. There were two classes of measurement: 1) objective, consisting of three forms, Mean Length of Response (MLR), Mean of Five Longest Responses (M5LR), and Type-Token Ratio (TTR); and 2) subjective, consisting of four seven-point rating scales representing a different aspect of language, structural sophistication, grammatical accuracy, creativity, and content. These subjective scales were taken from Elliot, Hirsh and Simmons (1967).

The speakers were two white male first grade children. They were considered to be of normal intelligence with no significant speech pathologies. The evaluators were instructed to listen to a tape recording of the two children and obtain a score for the several measures utilized. The results of the study did not indicate any significant biasing

effects for the four types of pre-information given to the evaluators. Also, no significant differences were revealed between the two speakers between the conditions of pre-information for the various classes of measurement (Manning and Beasley, 1971).

The authors provided several explanations as to the reason the speech pathologists were not biased. One was that masters' students were relatively resistent to evaluator bias. They are taught to form their own opinions about a client and not to do so before that client's language had been assessed by them personally. Another possibility could have been that the children were not seen personally by the evaluators, thus providing the opportunity for the evaluators to interact on a personal level with the children. Further, evaluators had no visual cues. Visual cues or auditory cues are independently sufficient to establish evaluator bias. However, studies have shown that a bias is more likely to occur when both visual and auditory cues are used simultaneously (Rosenthal and Fode, 1963; Adair and Epstein, 1967). If visual cues were employed, the likelihood of the occurrence of a biasing effect would have been increased.

If a subjective scale of measurement was employed that required the evaluator to indicate his own personal preference as to how he genuinely liked the speech of the child, there might have been a more accurate measure of how the evaluator truly felt about the child's language and speech. This rating scale could have been employed along with the

Elliot, Hirsh and Simmons (1967) scale. While the one set of scales could reflect upon the structural sophistication, grammatical accuracy, content, and creativity, the other scale could have been used for measuring the evaluators personal preference for the child's speech and language.

The semantic differential type scales devised by Osgood, Suci, and Tannanbaun (1957) may satisfy the above problem, since they reflect upon the listener's subjective reaction to the child's speech and language behavior. That is, they do not measure the stimulus per se, as do the Elliot et al. scales, but rather people's reactions to that stimulus.

Subjective measures in the case of the study by Manning and Beasley (1971) may be defined as descriptions of a speaker's language. The scale usually consists of five to seven points between two opposite terms. This type of scale was used by Elliot, Hirsh and Simmons (1967) in their investigation of the language of deaf children. These rating scales can be considered to coincide with the person-perception tasks mentioned on the preceeding pages. Rating scales can be considered to be subjective measures because they depend heavily on the rater's perceptions of the speaker's language (Manning and Beasley, 1971).

The possible effects of evaluator bias may be reflected in the amount of time devoted to the listening of the language sample by the evaluator. That is, individuals who spend more time listening to the language sample may in fact be more cautious in their evaluations. If

so, then those individuals would be less likely to reveal biasing effects. Related to this may be the evaluator's major professional area of interest. Thus, audiologists, who are trained to evaluate a different aspect of language behavior from speech pathologists, may very well reveal different results on an evaluator bias task than the speech pathologists.

Statement of the Problem

The physical and behavioral sciences have acknowledged and studied the various aspects of experimenter bias and experimenter effect. Investigators in audiology and speech pathology (Hipskind and Rintelmann, 1969; Nelson and Chaiklin, 1970; Williams, 1970; Meitus et al., 1972; Seigel, 1963; Manning and Beasley, 1971) have conducted studies involving the possibility of evaluator bias with regard to their respective professions. However, further research needs to be conducted to determine what possible effects, if any, evaluator bias has in the evaluation of the language skills of children. The purpose of this study is to determine whether an element of evaluator bias does exist in reference to subjective measures of a child's language performance with the aid of pre-biasing information and visual cues. Specifically, the following questions will be investigated:

1. Would speech pathologists and/or audiologists be influenced by information obtained prior to a language evaluation? Would the appearance of the child, as reflected by photographs, have any biasing effect on the evaluation of the child's language performance?

2. If the evaluator bias occurs, will it be more likely to occur because

the evaluator had a prior look at the child's picture and had a history of background information about the child or would it occur more readily if the evaluator had no prior information about the child but simply had a taped language sample to evaluate?

3. Would there be any major differences in the evaluations of the speech pathologists as opposed to the evaluations of the audiologists?

CHAPTER II

EXPERIMENTAL PROCEDURES

This study consisted of judgments of tape recordings of one speaker by thirty experimenters, fifteen audiologists and fifteen speech pathologists, who were randomly assigned to one of three conditions of pre-information, within each sub-group (speech pathologists and audiologists). Each condition was comprised of four forms of subjective measures of language. Each condition was represented by one of three types of pre-information intending to bias the experimenter. Visual cues were employed in the form of photographs.

Speaker

The speaker, used in an earlier study (Manning and Beasley, 1971), was a white male first grade child, age seven years-zero months, from an elementary school in Lansing, Michigan. The child was of normal intelligence according to school records and teacher's reports. The public school speech pathologist reported that his speech was not characterized by any significant misarticulations.

The speaker was asked to tell a story about each of the pictures in the Children's Apperception Test (Bellak, 1954). The story was recorded on a Panasonic Model RQ706S tape deck. The recording was obtained by a female speech pathologist, who was uninformed as to the

purpose of the study. Her instructions were to obtain a ten minute language sample from the speaker. The speaker and the speech pathologist were the only ones present while the recording was being taped.

Evaluators

The language evaluators were fifteen master's students in audiology and fifteen master's students in speech pathology from a large midwestern university. Before the evaluators listened to the tape recording of the speaker, they were asked to read information about the child contained in folders. Ten evaluators (five from each major) were given folders containing negative information, ten were given positive information folders to read, and ten were given a folder with no information.

Experimental Conditions

There were three conditions of pre-information used to bias the evaluators. The conditions were "negative information" (NI), "positive information" (PI), and "lack of information" (LI). The pre-information folders contained hypothetical "case reports" similar to ones found in a speech and hearing clinic or public school. The "lack of information" folder had only the child's age and a folder number. In the "negative information" and the "positive information" conditions, the folders included a folder number, age, and grade level. In addition to these factors, the negative and positive conditions also contained the social history of the child and his family, various measures of intelligence and achievement, and a photograph purported to be of the child in question.

The negative information folder contained a photograph of a child with gross malformations of the facial region. The positive information folder contained a photograph of a normal, "average looking" child. The lack of information folder contained neither picture nor information. The reports were placed in a manila folder and given to the evaluators just before they listened to a tape recording of the normal speaker. An example of the case report for the negative condition may be found in Appendix A-I, the positive condition Appendix A-II, and the lack of information condition Appendix A-III.

There were two scale classes of subjective measurement. The first class, used by Manning and Beasley (1971), was comprised of four seven-point rating scales. Each scale represented a different aspect of language. These scales were taken from Elliot, Hirsh and Simmons (1967) and included "structural sophistication," "grammatical accuracy," "creativity," and "content." (See Appendix B).

The second class of subjective measurement was comprised of five seven-point semantic differential type rating scales: like-dislike; mature-immature; good-bad; pleasant-unpleasant; and intelligible-unintelligible. (See Appendix C). These scales were of a more general nature than the first set of scales. Directions for rating the speaker indicated that the evaluator should listen very closely to the speaker and rate him as accurately as possible according to the stipulations on the rating sheets. The

evaluators were reminded that this would be their own personal judgment as to how the speaker's language and speech appealed to him.

The experiment was designed so that ten evaluators were randomly assigned to one of each of the three conditions of pre-information. The folders were coded by number as to the condition of pre-information they contained. The code was not familiar to the evaluator, thus minimizing the likelihood of the evaluator affecting the results.

Procedures

The evaluator was seated at a table in a small, quiet room. A tape recorder was placed on the table in front of the evaluator so that he or she could have easy access to it. The evaluator was instructed on how to use the tape recorder. The evaluator was then given the following printed instructions to read silently while the investigator read them orally:

> The purpose of this study is to obtain various ratings for each of four basic aspects of language and five measures as to the quality of language. You will receive instructions for obtaining these ratings now. You will hear a tape recording of a child's speech. Before listening to this tape recording you will be given some information to familiarize you with the child. Please read this carefully. After you have read the folder, you will listen to the tape recording for as long as you wish. After you have listened to the tape, please fill in the rating scales as accurately as possible. The purpose of the study in which you are participating requires that you do not discuss the folders or tape recording with anyone. Thank you for your cooperation.

The evaluator was allowed to ask further procedural questions if he so desired. Finally, a folder was given to the evaluator. The evaluator was allowed to read the folder for as long as he wished. After he read the folder, he listened to the tape. At this time, the evaluator was timed as to how long he listened to the tape recording. This time was recorded next to the evaluator's name and type of information given him to read. When the evaluator completed the rating scales, he was again reminded not to discuss the task with anyone.

Analysis

The investigator recorded all data (evaluator scores and time) by hand. These data are shown in Appendix D. The mean score for each type of pre-information in each scale were recorded and a standard deviation for each mean was also obtained.

CHAPTER III

RESULTS

Fifteen audiologists and fifteen speech pathologists were required to make subjective judgments of a tape recorded speech and language sample of a single child. There were two scale classes of subjective measurement. The first class was comprised of four seven-point rating scales. Each scale represented a different aspect of language. The scales were taken from Elliot et al. (1967). The second class of subjective measurement was comprised of five, seven-point rating scales. The evaluators were randomly assigned to one of three conditions of pre-information: negative information, positive information, and lack of information. Visual cues were employed in the form of photographs, intending to bias the evaluator.

The purpose of this study was to determine whether experimenter bias could exist in evaluating subjective measures of a child's language performance with the aid of pre-biasing information and visual cues. In addition, an attempt was made to determine whether one group (audiologists or speech pathologists) would be more influenced by biasing pre-information than the other.

Table 1 consists of mean scores and the standard deviations of these scores with regard to the Elliot et al., (1967) subjective language scales. Table 2 consists of standard deviations and mean scores of the five seven-point subjective scales. Table 3 depicts a mean listening duration and standard deviation in relation to the evaluator's major and condtion of pre-biasing information.

Language Scales

Overall effect of biasing pre-information.--Reference to Table 1 indicates that overall there were minimal differences between biasing conditions. However, there was a trend for the negative pre-information condition to be rated higher, i.e., better ($\bar{x}$=5.85), than the other two conditions. This is followed by the positive pre-information ($\bar{x}$=5.75) and lack of information ($\bar{x}$=5.65) conditions. There is a trend that suggests that audiologists were more influenced than speech pathologists. This influence is particularly evident for Creativity where audiologists rated the positive pre-information $\bar{x}$=6.40.

Effect of major and pre-information by scale.--Table 1 reveals the effect of pre-information according to the evaluator's major area of interest and type of pre-information. Overall, subjects tended to show minimal differences between classes of pre-information (mean scores range from 5.40 to 5.50) for the structural sophistication scale. Speech pathologists tended to rate the speaker slightly higher than did audiologists under all

three types of pre-information. However, between types of pre-information minimal differences existed.

Relative to grammatical accuracy, audiologists tended to evaluate the speaker highest with positive pre-information ($\bar{x}$=6.40), followed by negative pre-information ($\bar{x}$=5.80) and lack of information ($\bar{x}$=5.20), respectively. The speech pathologists rated the speaker for all pre-information conditions equally ($\bar{x}$=6.20).

Creativity scores were rated similarly for speech pathologists who had negative pre-information and lack of information folders ($\bar{x}$=6.20 and $\bar{x}$=6.00, respectively). The score for positive pre-information was lower ($\bar{x}$=5.20). Audiologists, however, tended to rate positive information higher than negative information, ($\bar{x}$=6.40 and $\bar{x}$=5.60, respectively). Lack of information was rated slightly lower than negative information ($\bar{x}$=5.20). Speech pathologists obtained equal scores for the positive pre-information and lack of information conditions for the Content scale ($\bar{x}$=5.80), while negative pre-information was rated higher ($\bar{x}$=6.00). The audiologists obtained the same mean score ($\bar{x}$=6.00) for negative pre-information as the speech pathologists. However, lack of information was rated lower ($\bar{x}$=5.80), followed by positive information ($\bar{x}$=5.20) for the audiologists. The standard deviation scores obtained for both speech pathologists and audiologists tended to be small. However, speech pathologists standard deviations were smaller than standard deviation scores for audiologists. These minimal differences in the

Table 1--Mean score data and standard deviations for each scale for the first subjective class of measurement (structural sophistication, grammatical accuracy, creativity and content) for audiologists, speech pathologists, and all subjects combined.

	STRUCTURAL SOPHISTICATION				GRAMMATICAL ACCURACY				CREATIVITY			
	NI	PI	LI	ALL	NI	PI	LI	ALL	NI	PI	LI	ALL
Audiologists												
Mean	5.00	5.20	5.20	5.13	5.80	6.40	5.20	5.80	5.60	6.40	5.20	5.73
Standard Deviation	1.41	0.84	1.30	1.13	1.30	0.55	0.84	1.01	1.70	0.55	1.80	1.44
Speech Pathologists												
Mean	6.00	5.60	5.80	5.80	6.20	6.20	6.20	6.20	6.20	5.20	6.00	5.80
Standard Deviation	0.71	1.14	0.84	0.86	0.84	0.84	0.84	1.04	0.45	0.84	1.00	0.86
All Subjects												
Mean	5.50	5.40	5.50	5.46	6.00	6.30	5.70	6.00	5.90	5.80	5.60	5.76
Standard Deviation	0.99	0.94	0.85	0.92	1.05	0.67	0.95	0.91	1.17	0.79	1.17	1.04

Table 1--Continued.

	CONTENT				COMBINED			
	NI	PI	LI	ALL	NI	PI	LI	ALL
Audiologists								
Mean	6.00	5.20	5.80	5.67	5.60	5.80	5.35	5.58
Standard Deviation	1.00	1.50	1.30	1.23	4.16	2.40	4.83	3.72
Speech Pathologists								
Mean	6.00	5.80	5.80	5.87	6.10	5.70	5.95	5.92
Standard Deviation	0.71	0.45	1.30	0.83	1.95	2.59	3.03	2.40
All Subjects								
Mean	6.00	5.50	5.80	5.76	5.85	5.75	5.65	5.75
Standard Deviation	0.82	1.08	1.23	1.04	3.24	2.21	3.95	3.14

standard deviation scores indicate that even though speech pathologists and audiologists differed slightly in separate aspects of each scale, their final ratings were generally consistent within the two groups.

The above findings substantiate the results of Manning and Beasley (1971) with regard to scores obtained for the speech pathologists, in that they do not appear to be influenced by pre-information. It appears, however, that audiologists tended to be influenced by such pre-information.

Descriptive Scales

Overall effect of biasing pre-information.--These descriptive scales included subjective measurement for five seven-point rating scales. The five aspects to be evaluated were like-dislike, mature-immature, good-bad, pleasant-unpleasant, and intelligible-unintelligible. The data for these scales are depicted in Table 2. These scales tend to result in larger standard deviation scores than the above language scales. In the case of Table 2 scores, the lower score indicates greater appeal of the speaker in relation to the five descriptive scales. Overall, evaluators gave the highest, i.e., poorer, ratings for negative pre-information ($\bar{x}$=2.68). Lack of information scores were the second highest with a mean rating of 2.20. Positive information scores were the lowest when all scores were combined. ($\bar{x}$=2.02).

Effect of major and pre-information by scale.--Table 2 reveals an evaluator effect pattern for audiologists and speech pathologists. Audiologists who received negative pre-information had rating scores the highest for

each aspect of the five scales (like-dislike $\bar{x}$=2.60; mature-immature $\bar{x}$=3.60; good-bad $\bar{x}$=2.40; pleasant-unpleasant $\bar{x}$=2.60; intelligible-unintelligible $\bar{x}$=2.80). Audiologists who had lack of information about the speaker had rating scores second highest for each aspect of the five scales; $\bar{x}$=1.80, $\bar{x}$=3.00, $\bar{x}$=2.20, $\bar{x}$=2.20, $\bar{x}$=2.00. (These figures are in the same order as the figures listed above.) Positive information scores were lowest, ranking third on all aspects of the rating scale ($\bar{x}$=1.40, $\bar{x}$=2.60, $\bar{x}$=1.80, $\bar{x}$=1.60, $\bar{x}$=1.40, respectively). Again, it can be stated that the audiologists tended to be more influenced by pre-information than were speech pathologists.

Speech pathologists gave similar results as those given by the audiologists for three of the five scales: mature-immature, $\bar{x}$ NI=2.80, $\bar{x}$ PI=2.20, $\bar{x}$ LI=2.40; good-bad, $\bar{x}$ NI=2.80, $\bar{x}$ PI=2.20, $\bar{x}$ LI=2.60; and pleasant-unpleasant, $\bar{x}$ NI=2.80, $\bar{x}$ PI=2.20, $\bar{x}$ LI=2.00. However, for the like-dislike scale the scores ranged as follows: $\bar{x}$ NI=2.60; $\bar{x}$ PI=2.80; $\bar{x}$ LI=2.00. Intelligible-unintelligible scores were approximately equal for the three types of information ($\bar{x}$ NI=1.80; $\bar{x}$ PI=2.00; $\bar{x}$ LI=1.80).

Generally, the standard deviation scores for speech pathologists tended to be slightly larger than did the standard deviation scores for audiologists. However, they were not of significance as to alter the consistent pattern.

Evaluator Listening Time

Table 3 indicates the means and standard deviations for the amount

Table 2--Mean score data and standard deviations for each scale in the second subjective class of measurement (like-dislike, mature-immature, good-bad, pleasant-unpleasant, intelligible-unintelligible) for audiologists, speech pathologists, and all subjects combined.

	LIKE-DISLIKE				MATURE-IMMATURE				GOOD-BAD			
	NI	PI	LI	ALL	NI	PI	LI	ALL	NI	PI	LI	ALL
Audiologists												
Mean	2.60	1.40	1.80	1.93	3.60	2.60	3.00	3.07	2.40	1.80	2.20	2.13
Standard Deviation	1.82	0.55	0.84	1.22	1.67	1.52	2.00	1.67	1.34	0.84	1.10	1.06
Speech Pathologists												
Mean	2.60	2.80	2.00	2.47	2.80	2.20	2.40	2.47	2.80	2.20	2.60	2.53
Standard Deviation	1.67	2.05	0.71	1.51	1.30	1.79	1.52	1.46	1.64	1.79	0.89	1.41
All Subjects												
Mean	2.60	2.10	1.90	2.20	3.20	2.40	2.70	2.76	2.60	2.00	2.40	2.33
Standard Deviation	1.65	1.60	0.74	1.37	1.48	1.58	1.70	1.57	1.43	1.33	0.97	2.57

Table 2--Continued.

	PLEASANT-UNPLEASANT				INTELLIGIBLE-UNINTELLIGIBLE				COMBINED			
	NI	PI	LI	ALL	NI	PI	LI	ALL	NI	PI	LI	ALL
Audiologists												
Mean	2.60	1.60	2.20	2.13	2.80	1.40	2.00	2.07	2.80	1.76	2.24	2.27
Standard Deviation	1.82	0.89	1.30	1.36	1.79	0.55	0	1.16	2.35	2.49	2.28	3.11
Speech Pathologists												
Mean	2.80	2.20	2.00	2.33	1.80	2.00	1.80	1.87	2.56	2.28	2.16	2.33
Standard Deviation	1.79	1.10	0	1.18	1.79	1.32	0.45	1.22	2.17	1.52	1.64	1.89
All Subjects												
Mean	2.70	1.90	2.10	2.23	2.30	1.70	1.90	1.96	2.68	2.02	2.20	2.30
Standard Deviation	1.70	0.99	0.88	1.25	1.77	2.05	0.32	1.17	1.96	2.38	1.89	2.53

of time each group of evaluators used for each type of pre-information. When the mean times were combined for all subjects, they were approximately of the same duration ($\bar{x}$ NI=7:43; $\bar{x}$ PI=7:26; $\bar{x}$ LI=7:55). However, when evaluating the audiologists' time factor as compared to the speech pathologists' time factor, it is revealed that the audiologists spent approximately two minutes less time listening to the speaker than did the speech pathologists ($\bar{x}$=6:42; $\bar{x}$=8:40, respectively). The most time spent listening in the audiology group was during the lack of information condition ($\bar{x}$=6:97), followed by the negative pre-information ($\bar{x}$=6:25) and positive pre-information ($\bar{x}$=6:05 respectively. However, the speech pathologists spent the most time listening during the negative pre-information condition ($\bar{x}$=8:61), followed by positive pre-information condition ($\bar{x}$=8:47), and lack of information condition ($\bar{x}$=8:13). It should be noted at this time that the speech pathologists went back over various sections of the tape-recording more often than did the audiologists.

Summary

This study investigated the existence of experimenter bias in subjective measures of a child's language performance under three conditions of pre-biasing information. The physical appearance of the child, as reflected by photographs, was utilized as a possible biasing factor in these evaluations. The type of pre-information, negative, positive, or lack of information, was considered in order to discover whether type of pre-information would differentially produce more bias in the evaluators.

Table 3--Mean time for listening to and evaluating taped speech and standard deviation for each condition of pre-information (negative, positive, and lack of information) for audiologists, speech pathologists and all subjects combined.

	NEGATIVE	POSITIVE INFORMATION	LACK OF INFORMATION	ALL
Audiologists				
Mean	6.25	6.05	6.97	6.42
Standard Deviation	1.77	1.82	2.12	1.81
Speech Pathologists				
Mean	8.61	8.47	8.13	8.40
Standard Deviation	1.82	1.82	1.82	2.65
All Subjects				
Mean	7.43	7.26	7.55	7.41
Standard Deviation	3.30	2.14	1.96	7.91

Finally, the scores obtained by speech pathologists were compared to scores obtained by audiologists to determine whether bias, on this task, would be more apt to appear in one professional group than the other.

These results suggest that pre-information, in general, played a minimal role in biasing the evaluators of the language sample in this study. There is a trend, however, for the audiologists to have been more influenced by the pre-information in a predicted direction than speech pathologists throughout both types of scaling procedures.

CHAPTER IV

DISCUSSION

Several investigators (Rosenthal, 1966, 1970; Evans and Rosenthal, 1969; and Rosenthal and Jacobson, 1968) have found the existence of experimenter bias in educational settings. Further, the controversial issue of "standard versus non-standard English," investigated by Williams (1970), suggested the possibility of bias with regard to speech pathologies in the speech of black children. In a study by Manning and Beasley (1971), experimenter bias was investigated as related to language evaluation by speech pathologists. They failed to induce a bias in their evaluators possibly because they only used an auditory stimulus.

The present study determined that biasing information, including visual cues, had minimal effect on the evaluation of language skills, utilizing the scales from Elliot et al. (1967), by either audiologists or speech pathologists. Scores obtained for the language scales reveal that the speech pathologists tended to rate the negative information with the highest mean score in three of the four condtions. These findings coincide with an earlier study (Manning and Beasley, 1971) where it was found that there was a tendency for the negative pre-information to lead to increased scores for two objective measures (Elliot et al., 1967).

However, Manning and Beasley (1971) also felt that the believability of the pre-information given to the experimenters was an important consideration. In their study, speaker #2 was rated higher for certain language scales under negative information. This speaker was also described as "immature." They felt that the difference between the negative information and the speaker's language abilities may have caused these abilities to be considered even better than they really were, i.e., a "reverse bias" phenomenon occurred. They feel that this could have been a cause for the speaker to be rated higher under negative information.

In the present study, speech pathologists rated the speaker highest for negative information for the language scales. The case history of this child revealed that he did not talk until age three and indicated that the child had a low IQ. The speech pathologists could have rated the speaker higher also for the same reason mentioned above by Manning and Beasley (1971).

The fact that speech pathologists and audiologists scores showed minimal differences for the language scales suggests that similar basic evaluative ideas exist between the two professions. One possible reason for the minimal differences could be based on the audiologist's previous training before concentrating on the field of audiology as a separate field. Essentially the audiologists were trained as were speech pathologists to consider all aspects of a child's speech and language when evaluating,

diagnosing, prognosing and carrying out actual therapy. This idea suggests that the relationships between the audiologist and speech pathologist should be very close including somewhat extensive interaction when evaluating and treating a client.

The results of the descriptive scales reveal a consistent effect in the mean scores of the audiologists. The negative information received the highest (i.e., poorest) score with respect to all conditions of the rating scale. The positive (i.e., best) pre-information received the lowest mean score for all conditions of the rating scale. These results correspond to the hypothesis that better scores would be predicted when pre-information was positive. Possibly, the audiologists tended to be more influenced by pre-information. When all subjects scores were combined, the mean scores and their standard deviations showed minimal differences for this group. Speech pathologists rated negative information highest on three of the five scales.

When all subjects' scores were combined, the mean scores for intelligible-unintelligible were the lowest, whereas the mean scores for mature-immature received the highest rating. This trend could suggest that these two variables have more influence when actually evaluating the language of a child in a clinical environment. Also, this trend could suggest that the type of pre-information had some influence on the scores.

Generally, the speech pathologists spent more time evaluating the speaker in all three types of pre-information than did the audiologists.

This result can be supported by the fact that speech pathologists are highly trained to evaluate the speech and language of an individual. They must take into consideration every aspect of an individual's background to determine any possible cause for his speech deficit. They are expected to take time in analyzing the person's speech in order to determine diagnosis and prognosis. The audiologist's main concern is with how efficiently a person perceives sound. The audiologist is highly trained to evaluate this particular aspect of an individual and likewise would be expected to spend a great deal of time in this evaluation. Thus, if the audiologists participated in an audiological study, they would be expected to spend more time than in a study involving speech pathology.

The failure of this study to induce a strong bias in the evaluators can be explained in several ways. The explanation given in the Manning and Beasley study (1971) may be offered as one alternative. They speculated that perhaps graduate students in speech pathology may be resistent to experimenter bias. Their training may teach them to form impressions of their own without the influence of other opinions. Meitus et al. (1972) provided the same speculation for their results. However, where Manning and Beasley felt this is an asset to the profession of speech pathology, Meitus et al. felt that clinicians should actually be influenced by case histories to form clinical impressions about a client. This investigator would tend to agree with the former author's opinion rather than the latter. The fact that the speech pathologist does not take into account the case

history before he evaluates the child would indicate a more accurate evaluation, in that the evaluator could evaluate the child more objectively without the influence of others.

Rosenthal and Jacobson (1968) discovered that the appearance of the child was one factor that affected gains made by the teacher's "intellectual bloomers." However, their study was conducted in an educational type atmosphere where the teacher was in direct contact, face to face, with the children. The direct contact plus the pre-biasing information may increase the likelihood of the occurrence of bias on the part of the teacher. The Meitus et al. study (1972) did not produce a bias when using the video tapes because although the clinicians could see the child in action, they still could not directly interact with the child. The authors do not place any emphasis on the usage of video tapes other than to state the conditions for which they were utilized in the procedures. The authors do not include this factor when discussing their analysis. They do not mention whether the videotape of the child with regards to his physical appearance, his behavior and the way he responds to speech stimuli have any influence one way or the other. The factor concerning the video tape should have been included in the analysis and discussed along with the case history variable; otherwise, it should not have been made a variable in the initial introduction.

The present study did not have interaction factors between clinician and client. This could be one explanation as to why the speech

pathologists and audiologists were not strongly biased when presented with visual cues.

Another factor involving direct interaction with the child could be the influence the experimenter has over the child he is evaluating. In the study by Beez (1968), the teachers who were given favorable expectations about their children tried to teach their children more symbols. The differences in teaching influences were clearly indicated. Rosenthal and Jacobson (1968) found that the children of whom the teachers were led to believe were "intellectual bloomers" did achieve far more than the other children. In the present study, the evaluator could not interact with the child; thus, his attitude could not be conveyed to the child to affect the child's performance.

Another factor which may have influenced the results is that there could have been the possibility that some of the evaluators knew the true nature of the study. This can be speculated because several of the clinicians asked the investigator whether the case histories were really true or whether the pictures were the real pictures of the speaker. Some evaluators also asked the investigator whether the study concerned experimenter bias. This being the case, the evaluators may have consciously or unconsciously given the speaker high scores in the case of the language scales and low scores in the case of the descriptive scales, thinking that this was what the investigator expected of them. This can be related to the concept of "demand characteristics." (Orne, 1962; 1969) Orne

suggested that the subject may feel that his performance on an experimental task should correspond with the results he feels the experimenter expects.

Still another factor that could have influenced the results was the fact that the experiment was conducted at the same university attended by the investigator. Information could have unintentionally been released from various sources connected with the investigator as well as the evaluators. Also, most of the evaluators were either friends or casual acquaintances of the investigator. This type of relationship could have caused the evaluators to try very hard to be "good" subjects. Orne (1962, 1969) and Rosenthal and Rosnow (1969) discuss the subject's intent on being a "good" subject. Orne stated that if the subject sees himself being evaluated while performing an experimental task, he will behave in such a manner as to please the investigator. A "good" subject may be many things: "to give the right response, to give the normal response, etcetera" (p. 145). This being the case, it would have been interesting to see what the results would have been if the investigator had no interaction with the evaluators prior to the experiment, that is, if the evaluators had been strangers to the investigator, or if the investigator was not present during the experimental task.

Clinical Implications

It would appear that this study indicates that speech pathologists and audiologists are relatively resistent to bias from pre-information. This

study supports the findings of Hipskind and Rintelmann (1969), Manning and Beasley (1971) and Meitus et al. (1972). It also underlines the caution of Manning and Beasley to continue to be resistent to such bias. This investigator cannot agree with Meitus et al. (1972) premise that the case history should have a biasing effect on the evaluator. It is felt that the clinician should form his own opinions regarding a client and these opinions should be set apart from any pre-information. It is felt that the evaluation of a child's speech and language would be much more accurate if it was based on the clinician's findings. This point should be emphasized especially when either the audiologist or speech pathologist has patients to evaluate and possibly admit for therapy. It should be remembered that most of the studies conducted where bias was found were conducted with a personal interaction between the experimenter and his subject (Rosenthal and Jacobson, 1968; Beez, 1968; Rosenthal, 1966, 1970). Research such as the present study should aid the speech pathologist and audiologist who will be faced with possible bias to be resistent to these biases when evaluating and diagnosing a client.

The results of the present study, in addition to the results of the studies by Manning and Beasley (1971), Meitus et al. (1972), Nelson and Chaiklin (1970) and Hipskind and Rintelmann (1969), indicate that speech pathologists and audiologists are not biased by case history information, therefore, at this point in time it can be safe to say that case history

information is not really regarded as essential information when evaluating a patient.

Implications for Future Research

Several factors may be considered for future studies concerning experimenter bias and its relationship to the professions of speech pathology and audiology. An important point to be investigated is the role of the speech pathologist in an audiological task. If a speech pathologist were placed in an audiological experiment, it could be speculated that he would have a tendency to be more influenced by pre-information. This is because the speech pathologist would be in a different environment from that accustomed to. It can further be speculated that the degree of influence would be greater for the speech pathologist than was found in the audiologist because the audiological task would probably not be too familiar to the speech pathologist. The speech pathologist would not have as much training experience with the audiological task as an audiologist would in speech pathology.

Another question to be answered is the effect of evaluator interaction with the speaker. If the evaluator worked directly with the child and used concrete test materials familiar to speech pathologists or audiologists, would this interaction have any effect considering the evaluator was given pre-information? In conjunction with evaluator interaction could be the consideration of the speaker's race. If the

child were of a minority race, would this have any influence on the evaluator to be more biased or less biased?

The type of pre-information used in the case history could be altered to include more matters concerning the child's intelligence and performance skills.

A further possibility for future research would be to include more experienced speech pathologists and audiologists. Would the individual who has considerable experience in his respective field evaluating, diagnosing and prognosing clients tend to be more influenced by pre-information or more resistent to pre-information?

A factor not previously considered would be including females in the group of speakers. Rosenthal (1966) found that the sex of the experimenter as well as the sex of the subject can influence the results in an experiment. The fact that males and females would be evaluating the speakers might make a difference in their ratings of females as opposed to males.

Hopefully, future research in the area of experimenter bias with regard to speech pathology and audiology will substantiate the results obtained by other studies (Manning and Beasley, 1971; Meitus et al., 1972; Hipskind and Rintelmann, 1969; Nelson and Chaiklin, 1970). This being the case, the results should serve as guidelines for people in other professions as well as speech pathology and audiology to

emphasize the unfavorable consequences that could arise as a result of evaluator bias.

LIST OF REFERENCES

Adair, J., and Epstein, J., Verbal cues in the mediation of experimenter bias. Psychological Reports, 22, 1045-1053 (1968).

Beez, W., Influence of biased psychological reports on teacher behavior and pupil performance. 76th Annual Convention of American Psychological Association, 3, 605-606 (1968).

Cordaro, L., and Ison, J., Observer bias in classical condtioning of the planarian, Psychological Reports, 13, 787-789 (1963).

Cowen, P., Weber, J., Hoddinott, B., and Klein, H., Mean length of spoken response as a function of stimulus, experimenter, and subject. Child Development, 38, 191-203 (1967).

Elliot, L., Hirsh, I., and Simmons, A., Language of young hearing impaired children. Language and Speech, 10, 141-158 (1967).

Evans, J., and Rosenthal, R., Interpersonal self-fulfilling prophecies: Further extrapolations from the laboratory to the classroom. 77th Annual Convention of American Psychological Association, 4, 371-372 (1969).

Hipskind, N., and Rintelmann, W., Effects of experimenter bias upon pure-tone and speech audiometry. Journal of Auditory Research, 9, 298-305 (1969).

Johnson, W., The Onset of Stuttering. Minneapolis: University of Minnesota Press (1959).

Johnson, W., Darley, F., and Spriestersbach, D., Diagnostic Methods in Speech Pathology. New York: Harper and Row (1952).

Manning, J., and Beasley, D., Experimenter bias and speech pathologists' evaluations of children's language skills. Paper presented to American Speech and Hearing Association, Chicago (1971).

Meitus, Irv J., Ringel, Robert L., House, Arthur S., and Hotchkiss, John C., Clinician bias in evaluating speech proficiency. In manuscript form, Purdue University (1972).

Merton, R., The self-fulfilling prophecy. Antioch Review, 8, 193-210 (1948).

Nelson, D., and Chaiklin, J., Writedown versus talkback scoring and scoring bias in speech discrimination testing. Journal of Speech and Hearing Research, 13, 645-654 (1970).

Orne, M., Demand characteristics and the concept of quasi-controls. In R. Rosenthal and R. Rosnow (Ed.) Artifact in Behavioral Research. New York: Academic Press (1969).

Orne, M., On the social psychology of the psychological experiment. American Psychologist, 17, 776-783 (1962).

Osgood, C., Suci, G., and Tannenbaum, P., The Measurement of Meaning, Illinois: University of Illinois Press (1957).

Rosenthal, R., Experimenter Effects in Behavioral Research, New York: Meredith Publishing Company (1966).

Rosenthal, R., and Fode, K., Psychology of the scientist: V three experiments in experimenter bias. Psychological Review, 12, 491-511 (1963).

Rosenthal, R., Teacher Expectations. Chapter to appear in G. S. Lesser (Ed.) Psychology and the Educational Process. New York: Scott Foresman (1970).

Rosenthal, R., and Fode, K., The effect of experimenter bias on the performance of the albino rat. Behavioral Science, 8, 183-189 (1963).

Rosenthal, R., and Jacobson, L., Pygmalion in the Classroom, New York: Holt, Rinehart and Winston (1968).

Rosenthal, R., Kohn, P., Greenfield, P., and Carota, N., Data desirability, experimenter expectancy, and the results of psychological research. Journal of Personality and Social Psychology, 3, 20-27 (1966).

Rosenthal, R., Persinger, G., and Fode, K., Experimenter bias, anxiety, and social desirability. Perceptual and Motor Skills, 15, 73-74 (1962).

Rosenthal, R., and Rosnow, R., The volunteer subject. In R. Rosenthal and R. Rosnow (Ed.) Artifact in Behavioral Research. New York: Academic Press (1969).

Siegel, G., Language behavior of adults and retarded children in interpersonal assemblies. Journal of Speech and Hearing Disorders Monograph Supplement, 32-53 (1963).

Siegel, G., Adult verbal behavior with retarded children labelled as "high" or "low" in verbal ability. American Journal of Mental Deficiency, 68, 417-424 (1963).

Williams, F., Language attitudes and social change. In F. Williams (Ed.) Language and Poverty. Chicago: Markham Publishing Company (1970).

APPENDIX A

CASE REPORTS

I. Negative Information

II. Positive Information

III. Lack of Information

CASE HISTORY RESUME

Folder #: II-20

Age: 7

Grade: 2

This child is the youngest of four children. His other brothers and sisters are in high school or college. His mother was in her early forties when he was born and his father was fifty-one. The pregnancy was normal but labor was difficult and prolonged. The child weighed four pounds at birth and was placed in an incubator immediately. It was determined that the child had a congenital heart defect. He did not leave the hospital until he was three months old and he has spent approximately two of his seven years in the hospital. He has had three operations on his heart so far and is currently in the hospital for his fourth operation.

When the parents were informed of the child's health problem, his mother decided to quit her job so she could spend more time with her child. She was constantly with the child during his long stays in the hospital. She would refuse to go home at night and stayed by the child's bedside. She would not let anyone else stay with the child alone. She refused to go out with her husband and began sleeping in the child's room. The other children were unable to have friends in the house and could not touch the baby. When the father suggested the mother was really crippling the child, the mother ordered him from the house. The parents are now divorced and the other children live with their father.

When the child last visited the hospital a psychiatrist spoke with his mother. The mother finally confessed extreme guilt feelings. She could not see how her attention could be harming her son, however.

The child did not talk until he was three years old. His responses still are very limited. A WISC test administered two weeks ago revealed a score of 73.

CASE HISTORY RESUME

Folder #: III-21

Age: 7

Grade: 2

This child is the youngest of four children. His other brothers and sisters are in high school or college. They all live at home. His mother was in her early forties when he was born and his father was 51. The pregnancy was normal but labor was difficult and prolonged. The child weighed four pounds at birth and was placed in an incubator immediately. It was determined that the child had a congenital heart defect. He did not leave the hospital until he was three months old and he has spent approximately two of his four years in a hospital. He has had three operations on his heart so far and is currently in the hospital for his fourth operation.

When the parents were informed of the child's health problem, his mother decided to quit her job as a remedial reading teacher so she could spend time with her child. She has always been with the child. The mother said that since the child would be handicapped physically she wanted him to have an active mind. Every toy bought for the child was educational. His mother read to him constantly and he was always surrounded by verbal stimulation. His brothers and sisters were very attentive and helpful. When the child was ten months, he was using meaningful words and was using two and three word combinations by fifteen months. He began recognizing printed words at three years and reading at about four.

A WISC administered two weeks ago revealed a score of 150.

CASE HISTORY RESUME

Folder #: IV-22

Age: 7

Grade: 2

APPENDIX B

LANGUAGE EVALUATION SCALE INCLUDING INSTRUCTIONS

Instructions[1]

When rating structural sophistication consider the maturity of the structure which includes: word order and amount of subordination and coordination, the utilization of prepositional phrases and subordinate clauses, as meaningful connections between words.

Structural Sophistication

simple ___ ___ ___ ___ ___ ___ ___ complex
1 2 3 4 5 6 7

When rating grammatical accuracy consider agreement of subjects and verbs, of pronouns and antecedents, of time words and tenses, and of inflectional endings. Consider also substitutions, omissions, errors and needless repetitions.

Grammatical Accuracy

poor ___ ___ ___ ___ ___ ___ ___ good
1 2 3 4 5 6 7

In rating creativity take into account originality, freshness and inventiveness of thought, surprise or other devices and vividness of the verbal picture.

Creativity

unimaginative ___ ___ ___ ___ ___ ___ ___ original
1 2 3 4 5 6 7

In rating content consider appropriateness and completeness of the description.

[1] Taken from Elliot, Hirsh and Simmons, "Language of Young Hearing Impaired Children," Language and Speech, 10, 1967, 141-158.

Content

sketchy	___	___	___	___	___	___	___	complete
	1	2	3	4	5	6	7	

APPENDIX C

DESCRIPTIVE SCALE

When rating the speaker, please consider the following criteria on the rating scales:

1. intelligible	___	___	___	___	___	___	___	unintelligible
	1	2	3	4	5	6	7	
2. good	___	___	___	___	___	___	___	bad
	1	2	3	4	5	6	7	
3. mature	___	___	___	___	___	___	___	immature
	1	2	3	4	5	6	7	
4. pleasant	___	___	___	___	___	___	___	unpleasant
	1	2	3	4	5	6	7	
5. like	___	___	___	___	___	___	___	dislike
	1	2	3	4	5	6	7	

APPRENDIX D

RAW DATA

NEGATIVE INFORMATION - SPEECH PATHOLOGISTS

Evaluator	Time	Structural Sophistication	Grammatical Accuracy	Creativity	Content
1	14.97	5	5	6	6
2	5.25	6	7	6	6
3	4.28	7	7	7	6
4	9.32	6	6	6	5
5	5.46	7	7	7	7
$\bar{x}$	8.61	6.00	6.20	6.20	5.80

NEGATIVE INFORMATION - AUDIOLOGISTS

1	9.09	4	6	6	6
2	4.93	6	5	7	7
3	6.85	4	4	5	5
4	5.46	7	7	7	7
5	4.92	4	7	3	5
$\bar{x}$	6.25	5.00	5.80	5.60	6.00

NEGATIVE INFORMATION - ALL SUBJECTS

$\bar{x}$	7.43	5.50	6.00	5.90	6.00

POSITIVE INFORMATION - SPEECH PATHOLOGISTS

Evaluator	Like-Dislike	Mature-Immature	Good-Bad	Pleasant-Unpleasant	Intelligible-Unintelligible
1	1	1	1	1	1
2	1	1	1	1	4
3	5	1	1	3	1
4	5	5	5	3	1
5	2	3	3	3	3
$\bar{x}$	2.80	2.20	2.20	2.20	2.00

POSITIVE INFORMATION - AUDIOLOGISTS

1	1	5	2	1	2
2	1	2	1	1	1
3	2	2	2	3	2
4	2	3	3	2	1
5	1	1	1	1	1
$\bar{x}$	1.40	2.60	1.80	1.60	1.40

POSITIVE INFORMATION - ALL SUBJECTS

$\bar{x}$	2.10	2.40	2.00	1.90	1.70

LACK OF INFORMATION - SPEECH PATHOLOGISTS

Evaluator	Time	Structural Sophistication	Grammatical Accuracy	Creativity	Content
1	9.89	5	7	5	4
2	7.30	6	6	5	5
3	8.78	6	6	7	7
4	9.31	5	5	6	6
5	5.37	7	7	7	7
$\bar{x}$	8.13	5.80	6.20	6.00	5.80

LACK OF INFORMATION - AUDIOLOGISTS

1	10.00	6	5	6	7
2	5.80	5	6	6	5
3	7.67	3	4	2	4
4	7.02	6	5	6	6
5	4.35	6	6	6	7
$\bar{x}$	6.97	5.20	5.20	5.20	5.80

LACK OF INFORMATION - ALL SUBJECTS

$\bar{x}$	7.55	5.50	5.70	5.60	5.80

LACK OF INFORMATION - SPEECH PATHOLOGISTS

Evaluator	Like-Dislike	Mature-Immature	Good-Bad	Pleasant-Unpleasant	Intelligible-Unintelligible
1	2	2	3	2	2
2	2	1	2	2	1
3	3	2	4	2	2
4	2	2	2	2	2
5	1	5	2	2	2
$\bar{x}$	2.00	2.40	2.60	2.00	1.80

LACK OF INFORMATION - AUDIOLOGISTS

1	2	1	1	1	2
2	3	4	4	4	2
3	1	6	2	1	2
4	1	2	2	2	2
5	2	2	2	3	2
$\bar{x}$	1.80	3.00	2.20	2.20	2.00

LACK OF INFORMATION - ALL SUBJECTS

$\bar{x}$	1.90	2.70	2.40	2.10	1.90