AN EXAMINATION OF THE HAWTHORNE EFFECT IN A
VERBAL LEARNING SITUATION IN AN
EDUCATIONAL SETTING

DISSERTATION

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Bert L. Simpson, B.A., M.A.
Denton, Texas
December, 1977

This study was an examination of the Hawthorne Effect in a verbal learning situation in an educational setting. The Hawthorne Effect was defined as the facilitating effect(s) produced in experimental situations when the subjects of the experiment expect that they are the objects of special attention. The purpose of the study was to determine if contamination by the Hawthorne Effect existed in an educational setting. Comparisons were made between "experimentally inexperienced" subjects and "experimentally experienced" subjects at both undergraduate and graduate levels. The task was to learn a list of paired-associate terms, and to show learning retention by immediate replication of those terms. The focus of the study was upon the expected differences in performance of the control and experimental groups produced as a result of an effort to persuade experimental subjects that they had "unique" characteristics which would cause them to be exceptionally proficient. The control groups were given the task by the course instructor in a usual classroom setting, as an example of a curriculum objective.
The population of this study was 394 undergraduate and graduate students enrolled in behavioral science and English classes in five North Texas area universities and colleges. A three-way analysis of variance technique was applied in an examination of three main variables. This technique allowed comparisons of the overall means and the two levels making up that factor. The three variables were (1) sex, (2) academic level, and (3) group. A one-way analysis of variance was applied to the data from the two academic disciplines, English and behavioral science.

The study found some support for the premise that the Hawthorne Effect may exist under specific rather than general conditions. Of the four experimental groups, the graduate English experimental group performed significantly better than the other groups. This is considered evidence supporting the Hawthorne Effect. Since no previous research in an educational setting has demonstrated this phenomenon, this finding has important implications.

The results showed mixed support for the four hypotheses. Hypothesis I stated that the experimental subjects would achieve a significantly greater mean score in the paired-associate learning task, as a measure of the Hawthorne Effect, than would control subjects. It was not supported. On further inspection of the collapsed data of the eight groups, it was observed that a number of
interactions had developed which masked a main effect interpre-
tation. Since the findings showed that these main variables were not acting independently, this general hypothesis was not supported.

Hypothesis II stated that English graduate subjects would achieve a significantly greater mean score on the paired-associate learning task as a measure of the Hawthorne Effect than would English undergraduates. It was supported. English graduate subjects did perform significantly better than English undergraduate subjects, with the English graduate, experimental group performing better than any other group.

Hypothesis III stated that behavioral science graduate subjects would achieve a significantly greater mean score on the paired-associate learning task as a measure of the Hawthorne Effect than would behavioral science undergraduates. It was not supported. No significant differences were found between the two groups.

Hypothesis IV stated that females would achieve a significantly greater mean score on the paired-associate task, as a measure of the Hawthorne Effect, than would males. This hypothesis was partially supported by the data. In behavioral science, female subjects did perform significantly better than male subjects, but significant differences in performance were not shown among male and
female subjects. With only partial support indicated, the research hypothesis was rejected.

Recommendations for further research were as follows: (1) the personality variables of the researcher and those of the subjects being tested should be thoroughly delineated; (2) sex differences in performance should be scrutinized further, since there were indications that females did react differently from males in the experimental situation; (3) other academic disciplines should be examined with respect to susceptibility to the Hawthorne Effect; and (4) other social settings should be examined with respect to the Hawthorne Effect contamination.
# TABLE OF CONTENTS

LIST OF TABLES ........................................... v

Chapter

I. INTRODUCTION ......................................... 1

Statement of the Problem
Purposes of the Study
Hypotheses
Basic Assumptions
Definitions of Terms
Research Design
Limitations of Design
The Significance of the Study
Chapter Bibliography

II. SYNTHESIS OF LITERATURE ......................... 11

Introduction
The Placebo Effect
The Halo Effect
The Teacher Expectancy Effect

Tutorial Studies of Teacher-Bias Effects
In-Classroom Studies of Teacher-Bias Effects
Evidence for a New Teacher-Expectancy Paradigm

The Hawthorne Effect

The Original Hawthorne Study
The Cook Study
Other Recent Studies

Summary
Chapter Bibliography

III. PROCEDURES FOR COLLECTION AND ANALYSIS
     OF DATA ............................................. 62

Research Implications for Design of the Present Study
Task Selection
LIST OF TABLES

Table                                    Page

  I. Analysis of Variance for Experimental 79
     and Control Groups

  II. Means and Standard Deviations for 79
       Experimental and Control Groups

  III. Analysis of Variance of English Subjects 79:
       For Sex, Academic Level and Group

  IV. Means and Standard Deviations for English 81
      Undergraduate and Graduate Subjects
      Relating to Academic Level

  V. Analysis of Variance for English Experi-
      mental Graduate and Undergraduate Subjects 82

  VI. Means and Standard Deviations for English 83
      Graduate and Undergraduate Experi-
      mental Subjects

  VII. Analysis of Variance for English Graduate 84
       Control and English Graduate
       Experimental Subjects

  VIII. Means and Standard Deviations for English 84
        Graduate Control and English Graduate
        Experimental Subjects

  IX. Analysis of Variance of Behavioral 85
      Science Subjects for Sex, Academic
      Level and Group

  X. Means and Standard Deviations for 86
      Behavioral Science Subjects by Sex
      and Group

  XI. Analysis of Variance for Male Subjects 88
      By Academic Level, Discipline and Group.

  XII. Means and Standard Deviations for Male 88
       Subjects in Behavioral Science and English

v
Table

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>XIII. Analysis of Variance for Female Subjects by Academic Level, Discipline and Group</td>
<td>89</td>
</tr>
<tr>
<td>XIV. Analysis of Variance for English Graduate and Undergraduate Female Subjects</td>
<td>90</td>
</tr>
<tr>
<td>XV. Means and Standard Deviations for English Graduate and Undergraduate Female Subjects</td>
<td>91</td>
</tr>
<tr>
<td>XVI. Analysis of Variance for English and Behavioral Science Female Subjects at Graduate Level</td>
<td>91</td>
</tr>
<tr>
<td>XVII. Means and Standard Deviations for English and Behavioral Science Graduate Female Subjects</td>
<td>92</td>
</tr>
<tr>
<td>XVIII. Analysis of Variance for Experimental English and Behavioral Science Female Subjects</td>
<td>92</td>
</tr>
<tr>
<td>XIX. Means and Standard Deviations for Experimental English and Behavioral Science Female Subjects</td>
<td>93</td>
</tr>
<tr>
<td>XX. Analysis of Variance for Behavioral Science Control and Experimental Female Subjects</td>
<td>94</td>
</tr>
<tr>
<td>XXI. Means and Standard Deviations for Behavioral Science Control and Experimental Subjects</td>
<td>94</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Since the classic studies conducted at the Hawthorne, Illinois, plant of Western Electric Company, a social phenomenon known as the "Hawthorne Effect" has been widely accepted by the scientific community. Even though the Hawthorne Effect was originally observed in an industrial setting, it has been assumed to be operating in other social settings, including the classroom. Simply stated, this concept refers to facilitating effects produced in experimental situations when the subjects of the experiment realize they are the objects of special attention.

Curiously, scant interest has been shown in replicating conditions to demonstrate the existence of this effect in a variety of settings. Apparently, the reason for such a paucity of research in this area is twofold: (1) the concept has immediate "face validity" in that it appears reasonable to assume that all human beings have a need to be given special attention; and (2) the concept as a term has not obtained reasonable consensus, since a survey of the literature would suggest that closely related phenomena are given various labels. For example, it has
been referred to as the "novelty effect," the "halo effect," attention, motivation and other similar terms (1).

Apart from the original interest of the Hawthorne researchers, who focused their attention on a human relations program for industry, there has been a subsequent concern in the social sciences as to the nature of a subject's reaction to experimental activities in relation to his individual needs. The efforts of Rosenthal (5) on experimenter bias, and Orne (3) in the area of subject's demand to have needs met in experimental situations appear to have particular relevance to the social/psychological variables which may be operating to produce the Hawthorne Effect.

In addition, Reiken (4) has also called attention to the need of subjects to please the experimenter. Reiken noted that subjects' attitudes and responses change as they become more aware of the nature of the experiment. A subject appears to be obligated to understand his involvement in the ongoing research, to have need to meet the experimenter's approval, and to have personal needs which have to be satisfied, each affecting his responses differently as the task progresses. Thus, the extent to which subjects have awareness of the foregoing variables should be a critical aspect of an experiment designed to determine if the Hawthorne Effect exists, and if so, under what conditions.
In examining the literature it was found that the most comprehensive study conducted in an educational setting was reported by Cook (2), with a fourth grade elementary school population. Cook manipulated a number of selected variables, i.e., curriculum material and teaching styles, in an effort to generate a Hawthorne Effect with respect to achievement gains, but no relationships were found. Consequently, since no study on the Hawthorne Effect using adult subjects has been reported, a study of this type seemed warranted.

It was suggested that the Hawthorne Effect might operate differentially in relation to academic levels, subject areas, and "awareness" variables. The present study utilized subjects in a university setting to investigate the possible existence of the effect among more mature and educationally more sophisticated students.

Statement of the Problem

The problem was an examination of the "Hawthorne Effect" in a verbal learning task situation in an educational setting.

Purpose of the Study

The purpose of the study was to determine if there was contamination caused by the "Hawthorne Effect" in an educational setting. Comparisons were made between "experimentally inexperienced" subjects in an English department of a
university at both undergraduate and graduate levels and "experimentally experienced" subjects in behavioral science courses at both undergraduate and graduate levels. The task was to learn a list of paired-associate terms and to show learning retention by immediate replication of those terms. The focus of the study was upon expected differences in performance produced as a result of an effort to persuade experimental subjects that they had "unique" characteristics which would cause them to be exceptionally proficient in this type of task. Variables examined were educational level, academic discipline, control versus experimental subjects, and sex.

Hypotheses

The following hypotheses were determined to be relevant to the purposes of this study:

I. The experimental subjects will achieve a significantly greater mean score on the paired-associate learning task than will the control subjects;

II. English graduates will achieve a significantly greater mean score on the paired-associate learning task as a measure of the Hawthorne Effect than will English undergraduates;

III. Behavioral Science graduates will achieve a significantly greater mean score on the paired-associate
learning task as a measure of the Hawthorne Effect than will behavioral science undergraduates;

IV. Females will achieve a significantly greater mean score on the paired-associate task as a measure of the Hawthorne Effect than will males.

Basic Assumptions

After orientation as to the purposes and objectives of the study, it was assumed that instructors of control groups (classes) handled the paired-associate task appropriately as a routine daily assignment in their classes.

Definition of Terms

Behavioral science--For this study, behavioral science means classes in education or psychology in which the subject content lends itself favorably to a verbal learning task assignment.

English--For this study, English means classes in which the subject content lends itself favorably to a paired-associate task assignment.

Graduate--For this study, selection of subjects was taken from officially designated graduate level courses in the above subject areas.

Undergraduate--For this study, selection of subjects was taken from courses officially designated undergraduate level in the above subject areas.
Subjects--For this study, subjects were chosen from classes which had the appropriate English and behavioral science designations. It is not implied that all these subjects were necessarily majoring in their respective disciplines, but were students who were enrolled in the classes participating in this study.

The Hawthorne Effect--For this study, the Hawthorne Effect refers to the facilitating effect produced in experimental situations when the subjects of the experiment realize they are the objects of special attention.

Research Design

A modified post-test-only, control-group model was used for the experimental design, since no pretest was necessary. All the experimental groups differed from control groups in that they were aware of their participation in a research study by direct instruction (see Appendix D, p. 114), i.e., direct cue. The control groups participated in the task under disguised conditions and every effort was made to prevent their awareness that this curriculum event was a departure from the standard classroom procedure (see Appendix C, p. 112). At the completion of the task, all subjects were given only the stimulus of the paired-associate term and were required to replicate the response.
Limitation of Design

Generalization may be limited to academic conditions similar to those in the sample. As earlier stated, it was suspected that the Hawthorne Effect did not exist in all academic situations; hence the effort to control for the variables mentioned.

The Significance of the Study

The present study was significant in that it provided definitive evidence of the possible existence of the Hawthorne Effect in undertaking cognitive tasks. Even though in recent years researchers have taken this phenomenon into account in planning educational studies, the evidence for it rests on the one classic study in an industrial setting. The most extensive educational study, by Cook (2), was not supportive.

After the Cook study, a number of important questions emerged. For instance, the cognitive development of the fourth grade sample used for the study would suggest a limitation with respect to possible awareness of their role in the study. An adult sample might indicate a greater tendency to establish purpose for and understanding of for their part in a particular activity. If so, it is possible that the Hawthorne Effect operates differently at different educational and developmental levels.
Another consideration was the difference in time span needed to show maximum effects. The Cook study was essentially longitudinal, covering a two-year time period, and it would appear that if the phenomenon appeared initially, the novelty of the experience would diminish over the two years. Additionally, other influencing variables, e.g., maturational factors, could be seriously mitigating the results.

Moreover, there may be sex differences related to the degree of susceptibility to this type of phenomenon. Females may show a greater need to establish meaning in behavioral situations. This notion seems consonant with female sex-role development, e.g., affiliation needs.

Finally, the degree of "awareness" may depend somewhat upon the experimental sophistication of the subjects. Those individuals who are accustomed to participation in experimental studies, e.g., behavioral science students, may be less susceptible to the Hawthorne Effect simply because they realize it is a routine assignment without any special meaning. However, those subjects who are ordinarily not involved in behavioral science research, e.g., English students, may be more likely to be affected by an unusual classroom activity, and consider their participation to have unusual significance.
Therefore, this study was deemed significant in that it would

1. Determine if a relationship existed between the Hawthorne Effect and differences in cognitive performance when using adults for subjects;

2. Determine if the Hawthorne Effect could be shown to occur in a short-term learning task situation;

3. Determine if sex was a factor in susceptibility to the Hawthorne Effect.
CHAPTER BIBLIOGRAPHY


2. __________, The Impact of the Hawthorne Effect in Experimental Designs in Educational Research, Columbus, Ohio, Ohio State University, United States Office of Education, 1967.


CHAPTER II

SYNTHESIS OF LITERATURE

In recent years a number of socially biasing variables have been predominently mentioned in the literature as having undesirable and/or contaminating influences on experimental results. Interestingly, these variables have been reported in specific situations, but they have been applied generally to a variety of settings. Of these factors the most frequently mentioned are the Teacher Expectancy Effect, the Halo Effect, the Placebo Effect, and the Hawthorne Effect. Whereas these variables may be critical in contributing to error in research findings, supporting evidence has been slow in appearing. Typically, if the phenomenon reported by some experimenter has an intuitive credibility, then it becomes popularized immediately. Unfortunately, in the case of most of these variables, replications have failed to show their existence, particularly in the general manner to which they have been applied. For example, even though the Hawthorne Effect was observed in an industrial setting, it has been assumed to apply in educational settings as well (39).

In reviewing the literature, it seems apparent that part of the difficulty in understanding these phenomena is
that conceptualization has not reached general consensus, causing research efforts to go into unproductive tangents. It appears unclear, for example, that any difference exists between the Novelty Effect and the novelty aspect seemingly inherent in the meaning attributed to the Hawthorne Effect. The Novelty Effect implies that a person may perform differently from his usual behavior in an unaccustomed situation. As a consequence, changes which may occur under temporary conditions are accepted as permanent gain. The Novelty Effect has been treated separately in the literature, apart from the other ingredients considered present in the Hawthorne Effect, and the result is confusion rather than clarification (59).

Another similar problem involves the Halo Effect and the Teacher Expectancy Effect. The Halo Effect implies an overall impression about a student held by a teacher which causes a bias in rating specific behaviors. In essence, this is also a description of one process in the development of the Teacher Expectancy phenomenon, which has been presumed to become potent enough as an influence to facilitate dramatic changes in student learning (76). The research on the Halo Effect, then, can be considered a special case, dealing specifically with rating scales as affected by certain teacher expectancies. In general, it may be said that the Halo Effect phenomenon has been extended and enhanced to be viewed in a larger context of Teacher Expectancy
to demonstrate how these developed teacher attitudes about students can also be affecting classroom student performance.

Further, "demand characteristics" has been researched independently by Orne (66) to explore the conditions under which students may acquiesce to the will of the experimenter in research participation. This research fits into the context of an enlarged paradigm for the Teacher Expectancy Effect suggested by Braun (12), which will be discussed in the following section.

Another problem apparent in analyzing the literature has been the failure of researchers to follow through a complete behavioral cycle in designing studies. As a consequence, it has become impossible to resolve some questions because of the incompleteness of information. A good example can be cited in probably the most researched of all the social phenomena mentioned, Teacher Expectancy, or the Self-Fulfilling Prophecy. In general, the many attempts to shed light on this perplexing but intriguing condition demonstrate the limitations of research efforts not only to show under what precise conditions teachers can be influenced by experimenters to change their opinions of students, but also under what precise conditions students may be influenced by their teacher to alter their performance in the classroom.

The two most complex concepts to be discussed in this review are the Teacher Expectancy Effect and the Hawthorne
Effect. There are philosophical similarities and contrasts which should be clarified.

The Teacher Expectancy Effect implies that attitudes of teachers toward students can be communicated in either subtle or obvious ways having the potency to cause dramatic change in the learning of their students. Thus, there is a dyadic interaction between a teacher and a student which has to be examined in light of the ingredients involved in order that whatever changes take place can be tied directly to the stimulus-response interchange. The past research on this matter has focused on the changes in the teachers' behavior, resulting from presumed acceptance of new information about some of his students. An analysis of these behavioral events has centered on the achievement gain of these students as a result of a presumed change in teacher behavior. There has been a simplistic assumption that teachers are passively accepting new information and acting upon it uniformly, as students are passively interpreting the influence of teacher behavior and as a consequence altering their learning style.

The Hawthorne Effect is a phenomenon presumed to occur in experimental situations as a result of students being aware they are participating as subjects in a study. As a consequence of feeling specially chosen, facilitating effects caused by the feeling will alter their performance irrespective of treatment variables.
These two notions, the Teacher Expectancy Effect and the Hawthorne Effect, have common conceptual components. The attempt in experimental studies to control the Hawthorne Effect has centered around the prevention of subjects from knowing that they were a part of an experimental situation in order that results would not be contaminated from awareness. As in the Teacher Expectancy conditions, there is a dyadic relationship involved, that of the experimenter and the subject(s). How subjects and experimenter react in this situation has been presumed to be without variation. Individual differences in response have been given only minimal attention.

Thus, these two phenomena are similar in that social biasing of behavioral outcomes is presumed to be taking place as a result of dyadic interaction. The Teacher Expectancy Effect and the Hawthorne Effect may both occur in the classroom as a result of presumed changes brought about within the student or the teacher, which alter his accustomed response. In both instances the information conveyed by the experimenter/teacher may be subtle and unintended and awareness of the dynamics operating may be low-keyed, but biasing nevertheless.

Certain focusing of these two phenomena has differed, however. In the Teacher Expectancy research, concern has been shown in examining how the teacher's behavior has affected certain student behaviors. In attempts to control
the Hawthorne Effect, more attention has been given to conditions affecting subject changes as a result of awareness differences caused by manipulation of direct and indirect cues. Thus, even though both phenomena have similar and dissimilar components within the total behavioral process, each reaches a biased outcome. Hence, it would seem the dyadic interchange in both instances should be under equal scrutiny.

Along with the social-biasing variables already discussed, another phenomenon noted in the literature is related to the concern of this study, the Placebo Effect. In brief, the Placebo Effect is an inert or irrelevant condition disguised to appear potent to subjects. The purpose is to demonstrate the subjects' susceptibility to persuasion. Since underlying all the phenomena mentioned is the attempt to discover the parameters of suggestibility in human behavior, the Placebo Effect fits into this framework. Commonly in behavioral research the Placebo Effect has become a technique used in research design for control purposes.

In the following sections, the Placebo Effect, the Halo Effect, the Teacher Expectancy Effect, and the Hawthorne Effect will be further detailed in a literature review.

The Placebo Effect

As a term, the Placebo Effect originated in the field of biochemical research. In order to avoid the Halo Effect, as
might be the case in a static-control group of an experimental design, a deceptive procedure is conceived. That is, all of the subjects are given pills, some of which are the drug being tested while others are chemically inactive. Any changes occurring in physical condition from the inert pills have been called the Placebo Effect. Simply, the changes which come about can be attributed to the belief of the patient that the inert drugs have the efficiency to cure. The degree to which subjects are susceptible to suggestion does appear to differ, though how or why some subjects are affected and others are not has not been clearly differentiated in the literature (42).

In applying the Placebo principle to behavioral science research, typically a placebo-control group is developed in which subjects are involved in an activity irrelevant to the purposes of the experiment. Meanwhile, in the static-control group no departure from the usual activities is allowed. If changes occur in the Placebo-Control group, then, it is assumed that these are due to the susceptibility of the subjects to suggestion only.

Whereas designing placebo groups as a measure of control in behavioral science has been somewhat popular in recent years, few studies have demonstrated that Placebo alone can account for the same results as treatment conditions (88). In none of an array of recent studies have the placebo groups shown greater effects than the
treatment groups. In all of the following educational experiments, some activity irrelevant to the purposes of the experiment was given to a control group in order to provide further support to the efficacy of the treatment as the cause for the results obtained. The results invariably supported the treatment applied. Studies were conducted by Boroughs (10), Boutwell (11), Dalis (22), Foley (35), Freiheit (38), Hughes (49), Kral (58), Peterson (70), Sulzen (91), and Wise (94).

A representative placebo study was undertaken by Klausmeier and Feldman (56), who were interested in the kinds of materials which could be used to best assist in concept attainment. Fourth grade children serving as experimental subjects were given lessons designed in the following manner: (1) a concept definition and placebo material; (2) a rational set of examples and non-examples of the concept and placebo material; (3) the definition, a rational set, and placebo material; or (4) the definition and three different rational sets. Control children read the placebo material only. The results showed that each experimental group performed significantly better than the control groups. The children reading a lesson with a definition and three rational sets performed significantly better than those who received only a definition. The authors concluded that the use of rational sets of concept instances and of a concept definition is validated as a "powerful controllable variable in instructional material."
In the above study, as in many others in behavioral science, the use of placebo in education has centered around the evaluation of innovative methods of teaching and/or curriculum application as opposed to those used conventionally. Even though in some instances a placebo group may show some improvement in performance over a static-control group, this could be attributed to a "novelty effect" being generated from participation in an unusual activity. As is indicated in the Klausmeier and Feldman study (56), the more closely the treatment materials were related to the intended purposes of the experiment, the more significant the gain.

From the research evidence presently available there does not seem to be sufficient gain realized in use of the placebo principle in the behavioral sciences. Whereas in a medical setting there have been instances of the potency of suggestion in the use of an inert pill as opposed to a bona fide drug, in behavioral science the use of an active but inappropriate control condition does not serve the same purpose. What is needed is an examination of the conditions under which suggestible influences can bias experimental results. Presently, the use of placebo groups in educational research seems to be a faddish technique without clear experimental support.
The Halo Effect

This widely acknowledged concept has been commonly observed in situations in which people are rated according to a criterion list of characteristics. The Halo Effect is the result of raters allowing an overall impression of the subject to influence their description of his specific behaviors. Thus, the Halo Effect can be positive or negative, and a variable factor in the determination of the nature of the data generated by an experiment. The obvious difficulty is that the rater's interpretation of the data is inconsistent, rating those with whom he has had some prior experience differently from those with whom he has had no acquaintance. In addition, certain psychological "sets" held by teachers cause them to view students according to these predispositions (50).

In a typical experiment to examine the phenomenon of bias in teachers' evaluation of pupils, Cahen (15) used as subjects a group of 256 senior and first-year graduate students, mostly female, working for an elementary teaching certificate.

The purposes of the study were twofold: (1) to see whether the behavior called halo or bias could be experimentally shown, and (2) to observe the relationship of selected personality, aptitude and background variables to halo scores. Two major hypotheses were developed: (1) that bias is a positive function of the amount of prior information
available to judges or scorers about the objects or persons
being rated; (2) that the amount of bias is greater when
early feedback of information or results is congruent with
the set held by the scorer or judge than when the early
feedback is incongruent with this set.

The experimental design consisted of having subjects
score a series of test booklets representing hypothetical
students. The four tests required somewhat subjective
scoring. Parallel tests representing high and low quality
pupil performances were created for the experiment. Prior
to scoring the booklets, the subjects were told about the
high statistical association between pupil aptitude and
pupil performance on the tests. Information about the
pupils' aptitudes was conveyed to the subject prior to
scoring and was intended to establish a set of expectancy
about student performance. One parallel form contained the
high quality set while the other parallel form contained
the low quality set. Bias was defined as the difference
between the scores assigned to the nominally parallel
forms. Subjects were randomly assigned to the nominally
parallel forms. Subjects were randomly assigned to sixteen
treatment combinations in two groups: (1) amount and kind
of information available to the scorer prior to scoring
the booklets, four levels; (2) logical orderings of scoring
of four booklets, based on the patterning of congruency
between the quality of set of expectancy conveyed to the scorer about the hypothetical child, and the actual pupil performance which was under experimental control.

The first hypothesis was confirmed: bias was related to the amount of information given to the subject prior to the scoring of the booklets. The second hypothesis did not reach conventional significance levels. However, the results for factor 2 were in the predicted direction.

In a more recent study, Jaeger and Freijo (50) examined race and sex as concomitants of composite halo in teachers' ratings of pupils. These ratings depended upon the race and sex of the teacher doing the rating and the race and sex of the pupil being rated. Two more specific assumptions were derived from the general hypothesis: (1) the degree of composite halo exhibited by teachers' ratings of pupils would be smaller, the smaller the teachers' perceived social distance from the pupils being rated, and (2) teachers' perceived social distance varies predictably as a function of the sex and race of the pupils being rated. An analysis of principal components and multidimensional scaling were used to investigate the hypotheses, using teachers' ratings of more than 8000 fourth grade pupils on twenty-one related behavior changes. The results confirmed the general hypothesis, but the secondary hypotheses based upon the second assumption, using functional relationships suggested by general sociological literature were not supported. The
authors suggest that the reason for this departure from the expected rating behavior may be due to this subculture's more sophisticated background and training, as differentiated from that of the general public.

Other representative studies supporting the general "halo" notion have been shown in various educational settings. Mitchell (64) conducted a study relating to ratings of institutional environments, postulating that students will form a "desirability halo" based on limited experience with an institution, causing a global orientation to be formed. A positive or negative halo effect is developed and acted upon when the student is required to make judgments about an environment on a desirability continuum. The High School Characteristics Index and a social desirability scale were administered to 2819 high school seniors in eleven high schools. The author concluded from the results that a "desirability halo" was developed, but appreciable wide differences were apparent in student bodies with respect to the strength and direction of that effect. Rubin (84), giving a general attitude questionnaire about reading to seventy-eight pupils in grades one and two, found that children of this level give answers in the "expected" direction as set by authority figures, illustrating the "halo effect."

Follman (36), in a study with college student ratings of teachers, showed that a generalized halo effect operated even though the instructions given requested that each item
be considered individually apart from all other items. Finally, Good and Brophy (45), investigating whether equal opportunities were given to boys and girls in a rural setting in four first grade classrooms, found that teachers did extend equal attention to each sex, but that boys generally received more negative comments during reading instructions. This finding was probably due to what the authors termed a generalized Halo Effect: that boys were more frequently disruptive than were girls.

Although all these results support the general halo notion, these research efforts suggest that all segments of the population do not have the same degrees of bias, nor do all members of the general population act uniformly upon this bias. It would appear that as social research becomes more sophisticated in its methodology and imagination, more differentiation will emerge.

The Halo Effect literature is likely to be viewed in the future as a special case or segment of the more complex behavior cycle encompassed in the teacher-expectancy notion. This literature can be incorporated into the general teacher-expectancy context to offer more specific information relating to the likely behavior variables of teachers and others as to the degree of their susceptibility to suggestion, and to more accurate prediction as to how the degree of assimilation of new information may be manifest in action.
The Teacher Expectancy Effect

Studies which eventually spawned the notions of Experimenter Bias and the Teacher Expectancy Effect came from a number of disparate sources. This body of research deals with changes in a subject's motivation brought about by attitudes and expectancies of the experimenter and, later, the teacher. These consist of (a) the prestige suggestion studies which tend to show that people will believe or do what prestigious sources suggest (47); (b) the "Hello-goodbye" effect in psychotherapy, showing that patients who merely have contact with a prestigious medical authority improve significantly over controls and almost as much as those who get prolonged therapy (37); (c) the goal-setting studies which show that setting goals for a person, particularly in the name of prestigious authorities like "science" or "research," improves performance (53, 63); (d) the parent-child interaction studies which show that parents who set higher standards of excellence for their sons are more likely to have sons with higher need for achievement (73). The common ingredient underlying this research seems to be that behavior change can be effected through the subject's perception that a respected authority figure expects him to perform in a prescribed manner. In other words, when subjects actually believe change can take place, the tendency appears to be that this attitude alone can be the motive facilitating the change.
The Experimenter-Bias Effect in psychological research has been thoroughly reviewed by Rosenthal (74, 75, 76, 77), Freidman (40), and Barber and Silver (5, 6). The Experimenter-Bias literature shows that in certain given situations experimenters may bias the performance of adults and children in psychological experiments (26, 27). It was this area of research which initially led Rosenthal and Jacobson (79) to extend this concept to the elementary classroom in an investigation of teacher-biasing effects on learning.

In the original studies conducted by Rosenthal and Jacobson (79), teachers were initially told that the authors were interested in further validation of an instrument designed to predict academic blooming or intellectual gain. A standardized intelligence test was administered to children in kindergarten through the fifth grade in May of the spring term. In the following September, an average of five children from each class were randomly designated as "spurters". As an experimental treatment the names of these "spurters" were casually given to the teachers. There was no real difference in the experimental subjects and the control subjects (all those children not designated as "spurters") except as developed in the minds of the teachers by the information given. Subsequent testing using the same children as subjects occurred four months later, at the end of the school year, and again in May of the following year. The authors predicted that if teachers expected achievement
gains in certain children, the behavior of teachers toward these children would cause a fulfillment of the prophecy.

The gains reported by the authors were mixed. Improvement was not uniform across grades. Gains at the end of the first year were the most dramatic for first and second grade children. At the end of the second year, the sixth grade children showed the greatest change. Verbal reports from teachers described children in the experimental group as being happier, more curious, more interesting, and as having a better chance of being successful in later life than control subjects.

The Rosenthal studies received immediate attention from both the scientific and popular media. Even though laymen and some educators began to accept the published findings as a revelation, the scientific community was skeptical. Snow (89) questioned the design and procedures of the study, the inadequate data analysis, and the lack of standardization of testing by teachers. Thorndike (93) and Snow (89) have also criticized the uses of tests of general ability for young and low-socio-economic-status children, making the credibility of the data doubtful. Further, Snow has questioned the sampling procedures of the study: the large attrition, the misleading graphs and tables, and the report of dramatic gains based on mean scores of the experimental group, whose pretest scores were below reasonable expectations for normal children. In reanalyzing the original data, Elashoff and Show (30) disputed the original claims of achievement gain noted for
grades three and six. Finally, a reanalysis of the first and second grade scores proved impossible, for teachers were unable to remember which students were the "bloomers."

Realizing that no research is likely to be completely free of error, Snow (89) suggested, however, that this study was far removed from the acceptable level of competence expected of respected investigators. Snow states, "one fears that the experimenters have convinced themselves, in the course of the analysis and of the book, that what they believed all along is true without further questions" (89, p. 199).

More recent efforts by Anderson and Rosenthal (3), Claiborn (17), Conn et al. (19), Evans and Rosenthal (31), Fleming and Anttonen (33), Flowers (34), Goldsmith and Fry (44), Jose (51), and Kester (55) have all been unsuccessful in replicating the Rosenthal-Jacobson findings. In a review of these materials, Baker and Crist (4) conclude that future research on this problem should be conceptualized as multivariate both in criterion measures and in teacher-pupil interaction.

**Tutorial Studies of Teacher-Bias Effects**

In a review of the literature related to this subject area, Dusek (25) groups studies into those tutorial situations in which expectancy efforts have been investigated. In an early study by Been (9), sixty graduate students in
education each taught a symbol learning task to one child in a summer Headstart program. Half the tutors were told that the child they were to teach was culturally deprived and would have trouble doing well in school. Another half of the tutors were told their children would do well in school in spite of their deprivation. Compared with the tutors of "non-problem" children, tutors of "problem" children tried to teach significantly fewer symbols (5.7 vs. 10.4), spent more time on non-teaching activities, rated the children lower in achievement, intellectual ability, and social competence and generally thought the task too difficult for the children. The children taught by these tutors did learn fewer symbols than the children who were expected to do well. These results suggested that teacher bias, which was induced by the experimenters, seemed to be translated into subtle teaching styles which affected the child's actual learning. However, rather than teacher bias, these data might be interpreted to mean that the tutors were simply trying to fit their teaching style to differing needs of students. Moreover, it is unlikely that ten minutes (the length of the study time) is enough time to gain a knowledgeable, well-informed opinion of a student's ability. The results may indicate only the tutor's understanding of the experimenter's statements about the students' abilities. These conditions affect the acceptance of whether any actual biasing occurred in the learning of the children.
Two studies by Robovits and Maehr (85, 86) dealt with tutors in a micro-teaching, racially mixed situation relating to students designated "gifted" and "non-gifted" by the experimenters. The studies revealed that tutors did not differ in the amount of attention given to the different categories of students, but did differ in the pattern of attention, requesting more information from the "gifted" and praising them more often. In the second experiment, subjects were mixed groups of "gifted" and "non-gifted" black and white children. There was a clear demonstration of differential teacher behavior as a function of expectancy and race. However, this difference may have been a reflection of the inexperience of the tutors, undergraduate college students.

Rothbart, Dalfen, and Barrett (80) conducted a study to investigate the relationship between teacher expectancy and changes in students' behavior. Subjects in the experiment were thirteen female college student trainees, each teaching an English literature lesson to four high school students, of whom two were designated as lacking in intellectual potential, and two were not. Although the tutors paid more attention to the "better" students, both groups were equally reinforced, negatively and positively. Although the data indicate that tutors treat children differently, it is still unclear what effect these treatments have on learning.
Obviously, the tutoring situation is unlike the real teacher-classroom context in a number of ways. Tutors are relatively inexperienced, do not know the teacher well, and are less mature than the typical teacher. However, it has been an accepted model for investigating the expectancy phenomenon.

Finally, other tutor studies have centered upon the personal characteristics of tutors. Carter (16), Panda and Guskin (67), and Rubovits and Maehr (86) all suggest that there may be certain teacher characteristics that predispose them to behave differently toward students with specific traits.

In summary, the results of the tutorial studies do not lend strong support to the expectancy notion. There is, however, a clear demonstration that tutors do respond differently to students designated as "gifted" and non-gifted." The children labeled as gifted seemed to have been involved in a more reinforcing interaction with the teacher, but the learning gain was not determined in most of the studies, and those interested in this aspect of the dyadic cycle failed to show that learning of the "gifted" increased. Moreover, any differences in teaching style attempted may have been the tutor's effort, as pointed out in the Brophy and Good studies (13), to adapt to what he perceived as differing student needs. Hence, it is doubtful that the tutoring situation can be generalized to that of professional
teachers, whose contact with children is far more pervasive. Thus, the tutorial studies should generally be considered as an adjunct to rather than identical with the typical situation.

**In-Classroom Studies of Teacher-Bias Effects**

Other than the Rosenthal studies on teacher bias, several important replications using the Rosenthal paradigm have been attempted in recent years.

Claiborn (17) conducted a study in an effort to demonstrate the Teacher-Expectancy Effect. Teachers were led to believe that certain students would show considerable intellectual growth during the year, and observations of student-teacher interaction in the classroom were also a part of the experimental procedure. The results indicated no difference in intellectual gain between experimental and control students. Additionally, as corroborated with a similar study of Alpert (2), there was no suggestion that the teachers responded differently toward children in the control and experimental groups.

Some differences in the Claiborn (17), and Rosenthal and Jacobson (79) designs should be noted. Whereas the biasing statements in the Rosenthal study were introduced at the start of the school year and were continued throughout the year, Claiborn initiated biasing information in the spring semester. It is possible that teacher attitudes had been formulated prior to this time of the year, and that the "set" was not easily subject to change. Further, only a two-month
interval between the pre and post testing was possible and it is likely that not enough time had passed to allow an effect to jell.

The most recent effort by Rosenthal et al. (78) to replicate his earlier findings was conducted in grades one through six of an all-black school. At the beginning of the school year the children were tested with forms of the TOGA (Test of General Ability), which was disguised as if to show creative potential. In addition, the students were asked to draw a man and to draw as many things as they could on one side of a sheet of paper. These latter two tests were used as a measure of creative ability. About 20 per cent of the children were selected as "bloomers" in creative ability and these names were given to teachers or those of students who would show unusual development in creative ability during the year. All tests were then repeated at the end of the school year, and the teacher-student interaction was recorded during the early and latter parts of the school year. The results for the school as a whole on the TOGA or creativity measures showed no evidence of teacher-expectancy effects. The exception was the fifth grade, in which the "bloomers" did score at a significantly higher level than did subjects in the control group on both the creativity and TOGA measures. Thus, Rosenthal et al. (78) concluded that teacher expectancy can influence test performance.

Dusek (25), however, states that the Rosenthal argument is not convincing: (1) the creativity test used had not been
standardized, and no evidence relating to its reliability and validity was available; (2) the scoring scheme, using the subjective judgment of eight creative persons has unknown reliability; (3) there is no explanation why the teacher-expectancy effects should occur only in the fifth grade, when previous research showed younger groups to be more affected (79). The study as a whole failed to replicate Rosenthal and Jacobson's earlier research.

An investigation conducted by Meichenbaum, Bowers, and Ross (60) showed positive teacher-expectancy results. The subjects were fourteen girls, six of whom were identified as potential intellectual bloomers. This information was given to six teachers who taught all fourteen girls. Both objective standardized tests and teacher-constructed and teacher-graded subjective tests were used. Analysis of data revealed that the potential bloomers scored higher than the control subjects on objective but not subjective tests of academic performance. Considerable teacher-behavior variation was also noted in interaction with the students. Some teachers increased positive interaction, while others decreased negative interactions, even though there was no change in total amount of interaction. This study seems to be the clearest demonstration yet published of teacher-bias effects in terms of students' performance and behavior as influenced by teachers' different behavior toward children.

Other recent studies of a similar nature by Jose and Cody (52)
and Mendels and Flanders (61) have not supported these findings.

A final body of research reported by Dusek and O'Connell (28), and O'Connell, Dusek and Wheeler (65) undertook to manipulate teacher expectancies regarding academic subjects and to measure the relationship between the teacher's own expectations regarding the student's performance and his achievement. At the beginning of the school year the Stanford Achievement Tests (SAT), disguised as tests to measure potential gains in language and arithmetic skills, were administered to two second and two fourth grade classes. At the same time teachers ranked students in class with respect to their own expectations of student performance in language and arithmetic skills at the end of the year. Then, teachers were given the names of eight children whose test scores indicated that they would show the greatest gain. This study was conducted over a period of one and one-half years. The results showed no evidence of teacher-bias effect on SAT performance, but teacher ranking was consistent and strongly related to SAT performance at each test occasion.

Brophy and Good (14) suggest that these findings would show that teachers do not bias students' performance but are good predictors of students' academic potential. Brophy and Good point out that teacher ranking is related to differential teacher-student interactions. Teachers demand better
performance from and are likely to praise students they rank higher. No relationship was assessed between teacher behavior and student performance.

In summary, it appears from the survey of the above studies that little support can be given to the notion that merely instructing teachers that certain students will show academic blooming is critical in affecting student performance. Only the Meichenbaum, Bowers, and Ross research (60) provides evidence for expectancy effects. However, this study included only sixteen subjects and cannot be considered typical of a classroom situation.

It seems evident that the research conducted in teacher-expectancy has not been fruitful within the Rosenthal paradigm, and it appears that the viability of the concept itself may be questioned. However, the intuitive acceptability of the notion is highly persuasive and a few of the studies reviewed lend some support.

Evidence for a New Teacher-Expectancy Paradigm

Three recent studies suggest a different approach should be taken to adequately interpret the Teacher-Expectancy phenomenon. Finn (32) notes that the research undertakings heretofore reported have been too simplistic. He holds that the interaction among learner variables and across learner-teacher variables warrants more substantive examination. To add further support, Kehle (54) developed elaborate descriptions of students, including a photograph, and asked teachers
to rate student personality characteristics and essay performance. The results showed that significant effects were attributed to the sex of the student, the sex-by-attractiveness interaction, and the sex-by-attractiveness-by-race-by intelligence interaction. The conclusions reached contend that complex interactions of student characteristics rather than simple variables in isolation bias teachers' perceptions and expectations of children. A similar study by Adams and Cohen (1) added further support to the multivariate composition of the expectancy concept.

In a recent and lengthy review of the expectancy material, Braun (12) explored the elaborate background experiences of the student which affect his perceptions of and responses to teacher behaviors, in relation to the perceptions of teachers about students based on their experiences and need systems. Braun reviewed studies pertaining to the different responses of teachers to their students with respect to physical appearance [Clifford and Walters (18), Dion (24), Kehle (54)], sex [Davidson and Lang (23), Meyer and Thompson (62), Pennock (68)], race [Howe (48), Rotter (81)], test results [Kehle (54), Sorotskin, Flemming and Anttonnen (90)], and knowledge of siblings [Seaver (87)]. Although studies concerning these variables were conducted independently of each other, they lend general support to the notion that all these factors interrelate in a myriad of ways to form a complex whole. The complexity of the various pattern
formations in behavioral manifestation is still known only at a primitive level.

In examining the other side of the relationship, some research attention has been given to teachers' experiences. The degree to which teachers may be susceptible to the influence of opinions contrary to their own views has been discussed by Gaite (41) and Dworkin (29). These authors conclude that information and personality are prime factors controlling the influence of expectancy cues. Thus, the degree to which imposed input data generates dissonance and the degree to which the teacher can reconcile this external information with what he has observed of the learner will probably determine the action he takes on the information presented.

The learner's responses to output cues from the teacher are just as complex to analyze as those of the teacher. However, externally induced variables generating feelings of inferiority-superiority have been shown to produce immediate effects of hostility, anger, distrust, and hopelessness in most children (69). Moreover, a study by Rosenthal, Baratz, and Hall (78) contends that the greatest gains in creativity resulted in classrooms where teachers were judged to be likeable, interested in children, enthusiastic, professional and inclined to give encouragement.

Another approach to understanding a child's responses to his environment is to study his developing self-concept.
Gergen (43) supports the notion that the number of confirmations a child receives regarding his self-image and the consistency of these confirmations influence change. The potency of an expectancy cue would depend upon the ability of the self-image to withstand dissonance brought about by conflicting material and the perceived credibility of the source.

Thus, Braun concludes that "there is a close relationship between the teacher's expectations for the learner, the teacher's treatment (output) of the learner and ultimately, the child's self expectation" (12, p. 208).

Braun's solution to the teacher-expectancy problem has a high degree of credibility even though the issue becomes more complex. It is within this framework that progress is likely to be made in the future.

The Hawthorne Effect

Of the social-biasing variables discussed in this chapter, the two most complex in scope appear to be the Teacher-Expectancy Effect and the Hawthorne Effect. Whereas a large quantity of research material has been generated from the Teacher-Expectancy notion, the Hawthorne Effect has had the opposite influence. In an exhaustive literature search, only a few attempts have been found which focus directly on the phenomenon. Whereas research designs have regularly sought to control for its
possible existence, lack of interest in developing scientific support has been noticeable.

Although the condition may be difficult to translate into a research design, this may not account fully for the neglect. This notion appears to serve a useful purpose in that unanticipated results in experiments may be explained by attributing the results to this uncontrolled variable. Possibly, this concept has served as a convenience for experimenters who had errors of control or design, or both, and who wished to disguise the real difficulty.

Generally, the Hawthorne Effect centers in the area of the subject's cognitive activity, particularly his perception of his immediate environment. As a concept, the Hawthorne Effect incorporates aspects of the contaminating social variables previously discussed. The bias created appears to be influenced by expectancy notions, demand characteristics, novelty aspects in an altered social structure, and level of awareness of what purposes are intended in a given situation. Thus the biasing takes place primarily within the province of the subject's activity, according to the traditional paradigm.

One of the dangers in generalizing from the original industrial setting to the classroom is the difference in the cognitive sophistication of the subject. The subjects in the original Hawthorne experiment were six adult females whose expectations were related to the cognitive limits
imposed by a routine industrial task. What constitutes
total "awareness" of the research purposes is speculative.
In general, adult students in an educational setting probably
operate at a more sophisticated cognitive level and "aware-
ness" may be presumed to be at a higher cognitive level.
Two other considerations are important. Subjects used in
all the Hawthorne studies conducted in an educational setting
were taken from elementary and secondary classes. No studies
have been found which used adult subjects in an educational
setting. Major differences exist between adults and children
in cognitive functioning, particularly in abstract abilities.
It is curious that replications have not been concerned with
this validity problem. In addition, Thorndike (92) pointed
out that the novelty factor which should be considered a
central component of the Hawthorne effect will decay with
the passage of time. Hence this appears to be a serious
criticism of the Cook study (21) which was longitudinal in
scope.

The "Novelty" Concept

As has been noted, the novelty concept has been often
related to or used synonymously with the Hawthorne Effect,
although at times it has been treated independently in the
literature. The observation that a subject may perform
differently from his usual behavior in an unaccustomed
situation has received some attention from experimenters.
The term implies that results are biased in that changes which may occur under temporary conditions are accepted as a permanent gain. The novelty of the situation has temporal characteristics and "real" gains may disappear when the novelty becomes routine.

Several studies illustrate how the phenomenon has been examined. Popham (71), for example, was concerned with the influence of the novelty effect upon teaching machine learning and more conventional teaching methods, and found that differences noted initially disappeared within a year. Beck (8), in assessing some newly designed reading materials at the first grade level, found that the novelty variable was subsumed to other variables such as teacher behavior, pupil differences in reading readiness, and parental assistance. In a study by Linder and Whitehurst (59), an attempt was made to shed light on the question of whether the positive student attitudes toward personalized instruction (courses in which students move at their own rate of speed) reported in recent research can be explained in terms of a novelty effect. A departmental attitude survey was sent to all undergraduate students taking personalized instruction psychology courses at C. W. Post College. A correlational analysis revealed no significant relationship between student attitudes and number of personalized psychology courses taken concurrently. These data do not support the "novelty effect" hypothesis.
In a discussion of the original Hawthorne studies, in the following section, the novelty aspect of the experimental setting is presumed to be of major consequence in creating the effect.

The Original Hawthorne Study

The most frequently cited report of the original Hawthorne study is that of Roethisberger and Dickson (72). The intent of the study was to examine the relationship between working conditions and productivity. The subjects were a team of female telephone relay assemblers who were put into a room where the physical environment could be controlled. A rigorous attempt was made to control variables influencing production—humidity, temperature, lighting, rest periods, and others. The variables were manipulated in order to obtain the optimal combination of working conditions to achieve maximum output. As the study progressed, with each improvement in lighting or temperature, production increased. Initially, it appeared that improved working conditions improved production. However, in investigating production records, it was revealed that when one of a group of five female employees was absent, instead of a reduction in output, as would be expected, production actually increased. Puzzled by this situation, the investigators reversed the experimental conditions, eliminating rest periods, lengthening the work period, and reversing the improvements in illumination
and humidity. Despite these adverse conditions, production continued at the same level. It was found that the variable controlling the unexpected results was the feeling of the subjects that they were being given special attention, and the novelty of being associated with the experiment provided an interest and appeal which their work generally lacked. Hence they responded positively no matter how the variables were changed. Presumably, the experiment suggested the illusive nature of motivation and its effect on behavior.

Initially, little interest was shown in generalizing this phenomena to other types of experimental settings. The term "Hawthorne Effect" did not appear until French (39) suggested that in educational experimentation, field settings were ordinarily less artificial than were laboratory conditions, except for situations which tended to create awareness in subjects of special treatment.

Since the focus of the present research was directed toward the Hawthorne Effect, one other important comment should be made in light of recent information released subsequent to this present investigation. Horne (46) presented some further details concerning the original Hawthorne Study, drawn from an examination of original reports from conversations with several researchers who had previously studied the experiments and from a man who had been a supervisor in the test room.
From these sources emerged two important facts hitherto unreported. (1) The record of how many relays each worker produced was available for every half hour and at the end of each day. Thus, they were given feedback on the day's production. (2) At the beginning of the experiment, the women who were later subjects were being paid according to the output of the entire large department, which was not affected very much by individual effort. When the experiment began they were switched to a piece-work system under which their wages were based on the output of their six-woman group.

Thus, it would appear that these two unreported facts may explain the difficulty later researchers have had in demonstrating the effect. Since the higher the production the larger would be the monetary reward, reinforcement would account for the increased production regardless of the particular treatment variables being examined. Obviously then, the Hawthorne Effect did not exist in the manner suggested by the original researchers.

As conceptualized, the present study followed the traditional paradigm, but attempted to go further than past research in examination of such variables as sex differences, college students as subjects, two dissimilar academic disciplines and two academic levels. In light of this most
recent information, the Hawthorne Effect will require some rethinking.

The Cook Study

After the original Hawthorne studies, the next direct investigation was conducted by Cook (21) in an exhaustive study which lasted over a period of two years. This undertaking was funded with a grant from the U. S. Office of Education and sought to answer three general questions: (1) Can the Hawthorne Effect be identified as the variable accounting for all or a major portion of the difference or lack of difference between experimental and control groups in educational research studies? (2) Can the component elements of the Hawthorne Effect be distinguished? (3) Can control measures be established that are efficient and effective?

Two methods of collecting data were utilized: field research and literature analysis. The field experimentation was conducted with sixteen fourth grade classes and continuing with the same population into the fifth grade. Cook manipulated one direct cue and two indirect cues. The direct cue was the instruction to some of the students that they were participating in an experiment. The indirect cues were the introduction of a School Mathematics Study Group Program and a different teacher for the mathematics instruction. The dependent variable was the gain in student achievement.
Cook (20) analyzed three hundred fifty research studies which mentioned the Hawthorne Effect or synonyms. The literature revealed that individuals writing about the Hawthorne Effect were not in agreement as to the meaning or definition of the concept. Actual control procedures were not related to results in any systematic way. Control mechanisms were selected on the basis of intuition. There was no empirical evidence that these procedures actually controlled the Hawthorne Effect.

The field research study revealed no significant differences in student gain in mathematics achieved at the end of the first year, second year, and two-year period as a result of the main effects of direct and indirect cues. In addition, no significant relationships were observed between measured awareness of participation and gains in mathematics at the end of the first year, second year and the two-year period combined.

Other Recent Studies

Bauernfeind and Olson (7) reported the results of a field experimentation with the Hawthorne Effect. They investigated the influence of two components of the Hawthorne Effect: (1) direct cue--the explanation to students that they were participating in an experiment; and (2) indirect cue--the use of special test administrators.

The major study was conducted with students in grades three, five, seven, and nine in a culturally heterogeneous
urban community. The number of intact classroom units per grade ranged from twenty-three to twenty-seven. A replication study included students in grades four, six, and eight in a large rural school district. The number of classroom units per grade in the replication study ranged from four to seven.

The experimental task involved the administration of two short tests of reading and arithmetic within a ninety-minute time period. Olson concluded that the Hawthorne Effect was (1) not in existence for pupils of these ages in this particular study, or (2) not evoked by the cues utilized in this experiment, or (3) not sufficiently potent to produce significant changes.

In another study Kleinberg (57) was interested in the novelty aspect of the Hawthorne Effect as related to improvement of reading skills through use of the tachistoscope. The experimenter sought to show differences in reading achievement by comparing the tachistoscope with other machine and non-machine methods. The population was five fourth grade classes totaling 163 pupils. The Gates-Mac-Ginitie Reading Test was administered to all pupils as a pre-test and classes were evaluated on the mean, standard deviation, and general range of scores. The five classes were then randomly assigned to one of the three treatment groups. The results supported the hypothesis that the tachistoscope was no more effective than use of other types of mechanical equipment. There was no clear evidence that the Hawthorne Effect was necessarily operating.
Rubeck (83) examined the relationship of the Hawthorne Effect in producing change in reading test performance as a result of pseudo-experimental manipulation and timed and untimed tests. The four treatment groups consisted of (1) experimental timed, (2) experimental untimed, (3) control timed, and (4) control untimed. The dependent variable for the study was the post-test score on the Nelson/Denny Reading Test adjusted for pre-test differences among experimental and control groups. Twenty-four classrooms from grades three through eight were selected for the research sample. The groups were given the pre-test under the timed and untimed conditions. Then, experimental groups were subject to activities which had no known relationship to reading and given a post-test. The results showed no significant difference in groups, and therefore no perceptible Hawthorne Effect was noted.

In a later article, Rubeck (82) summarized the research findings on the Hawthorne Effect, stating that it has not been demonstrated to be a confounder of educational research findings. The author calls for more systematic attention from investigators to further clarify the issues involved. Particularly does the author call attention to the possibility that the personality of the investigator was an influencing factor in all these experiments. A final suggestion calls for a more systematic manipulation of time.
Summary

The review of significant studies relating to the social-biasing variables, the Placebo Effect, the Halo Effect, the Teacher Expectancy Effect, and the Hawthorne Effect indicated limited progress has been made in understanding their complex properties.

The Placebo Effect has been used by researchers to modify the control condition in an experiment by designing an action-control group to engage in a presumably inert activity in relation to the purposes of the treatment. The intended object is to determine if activity alone has any potency in causing the same effect as the treatment. Even though the Placebo Effect was originally shown under certain circumstances in a medical setting, it seemingly has not served a very useful purpose in education. In the educational studies reviewed, no dramatic effect was produced by a placebo control group. Indeed, Gephart and Antonoplos (42) consider the use of the placebo a contributor to error through inducing altered performance by control subjects, resulting in an inaccurate comparison between groups. Consequently, it would appear that in education, using a placebo-control group may be counterproductive.

Although the Halo Effect and the Teacher Expectancy Effect have some common conceptual notions, they have been studied separately. In essence, the Halo Effect is a fragment of the Teacher Expectancy Effect. Since the Halo
Effect implies that a specific impression is generalized to an overall estimate of one's abilities, this phenomenon can be considered the first portion of the Teacher Expectancy notion. The Halo Effect, being a more simple construct, shows that a specific student quality, which causes an attitude to be developed in a teacher, can affect the teacher's behavior by precipitating a generalization of the bias in rating all abilities. While this phenomenon has been abundantly supported by research, the effect of a bias in changing the behavior of students (the Teacher Expectancy Effect) has not been clearly supported by research. It has been suggested by Braun (12) that within a new research paradigm the Teacher Expectancy Effect may survive and may be supported finally by research evidence.

The Hawthorne Effect presents some of the same difficulties as the Teacher Expectancy Effect in that it has intuitive credibility, but has not been supported by educational research. Indeed, recent material by Horn (46) questions the validity of the concept as reported in the original industrial study. It would appear that in the original study, monetary reward through accelerated productivity may have been more motivating than the participants' feeling that they were the objects of special attention. The illusive nature of the "special attention feeling" may be of sufficient magnitude to force a general abandonment of the concept as unresearchable in the context of the traditional
paradigm. Thus, as an alternative, it may be necessary to consider personality variables of the researcher and the subjects as potent in the possible retention of the concept.
CHAPTER BIBLIOGRAPHY


22. Dalis, Gus T., Using Precise Objectives to Enhance Student Achievement in Health Education, report to Los Angeles County Superintendent of Schools, Los Angeles, California, 1970.


59. Linder, Steve and Carol Whitehurst, "Is There a Novelty Effect on Student Attitudes Toward Personalized Instruction?" *Journal of Experimental Education*, 42 (Fall, 1973), 42-44.


68. Pennock, Clifford D., "Student Teacher Expectations for Primary Level Boys' Reading Achievement," Dissertation Abstracts International, 32 (1972), 5657-A.


CHAPTER III

PROCEDURES FOR COLLECTION AND ANALYSIS OF DATA

In relation to previous Hawthorne Effect research designs, the present study followed the traditional paradigm structure. Additionally the study also attempted to examine variables not previously considered by researchers but which appear to be relevant. As has been noted, the nature of the Hawthorne Effect has probably discouraged researchers from directly investigating the phenomenon. The observed effect was interpreted to mean that people can be motivated to perform at a higher base level than normal for them, as a result of feeling that they are objects of special attention. This intuitively credible premise has apparently hindered rather than generated supporting research. In addition, it is difficult to design conditions in the classroom which would clearly indicate the existence of the Hawthorne Effect. The most perplexing problem is to have reasonable confidence that the experimental conditions will actually cause all subjects to feel that they are objects of special attention, and hence to display a higher level of performance. Nevertheless, in spite of the paucity of evidence, some researchers have found the Hawthorne Effect a convenient means to account for unexpected results in experiments and have willingly embraced the concept.

62
Investigations in education subsequent to the original industrial study have involved elementary-level subjects in a treatment program related to achievement gain. In these experiments, innovative curricula, i.e., reading, mathematics, were given to designated experimental and control groups subsequent to being pretested by standard achievement tests. In experimental groups, typically, direct and indirect cues were manipulated. In the direct-cue conditions, subjects were informed that they were participating in an experiment and an effort was made to make them feel that they were specially chosen. In indirect-cue conditions, different instructions from those customary or an abrupt change in instructors might serve to alert students that they were participating in an experiment. When experimental and control groups are compared with respect to achievement gain on the posttest, if the treatment group is significantly higher in gain, the Hawthorne Effect would presumably be operating. In all studies reviewed none have demonstrated the Hawthorne Effect.

Research Implications for Design of the Present Study

The educational studies reviewed have differed in a number of respects from the original Hawthorne Study. (1) Subjects in the industrial study were adult females, whereas subjects in all the educational studies have been elementary age children. (2) All educational studies have analyzed results
without differentiating a possible sex variable, whereas the original study included only adult female subjects. (3) In all educational studies no particular effort was made to manipulate the environment as in the original industrial study: that is, subjects were not actually moved from one setting to another to dramatize that they were being treated differently, whereas in the industrial study, all subjects were moved into a special room, apart from the other factory workers.

Thus one of the purposes of the present educational study was to incorporate into the design aspects of the original industrial study which have been overlooked. Hence a number of assumptions were made. Since adults were used in the original study, adults were used in this educational study. Attempts were made in this study to examine possible sex differences in performance, since previous research had not examined this variable. Finally, in the present study, experimental groups were moved from their accustomed setting into an unaccustomed setting.

In order to use adults as subjects in an educational setting, a college population was determined to be appropriate. It was speculated that susceptibility to suggestion (or to the Hawthorne Effect) may depend upon the opportunities students have had to be a part of experimental situations. Students in academic disciplines which have frequent opportunities to participate in or be acquainted with research
studies, e.g., social science, may suspect experimental deception. Consequently, students in two contrasting disciplines should represent experienced versus inexperienced experimental subjects, e.g., behavioral sciences versus English. Thus it is proposed that some students may be more likely to demonstrate the Hawthorne Effect than others, and a more careful scrutiny of differences is needed.

Finally, students may vary according to academic level. It is suspected that graduate students may be more acquiescing to classroom tasks than are undergraduates, since the former should be more seriously concerned with variables which might affect their professional standing, e.g., teacher approval. If these assumptions are valid, then, the Hawthorne Effect should be operating more dramatically at the graduate rather than undergraduate level. Hence, the need existed to use both graduate and undergraduate subjects in the study.

Task Selection

A paired-associate task was selected as an appropriate vehicle to use in researching the Hawthorne Effect in an educational setting. There were several reasons for this choice: (1) Since the nonsense-syllable instrument has been used traditionally as a means of examining the processes involved in learning, it is an excellent means for equating groups with respect to learning. That is, since the instrument precludes prior learning, all groups start with the
same learning base. In a paired-associate design, a tri-gram, containing three nonsense letters, is attached to an adjective. The learner is required to develop a means of associating the appropriate tri-gram adjective relationship for later recall. (2) The paired-associate task is representative of a cognitive exercise which can be generalized to others in the typical academic situation, i.e., general associative learning processes. (3) The paired-associate task is an appropriate verbal exercise which would fit into the traditional curriculum of at least two academic disciplines, i.e., behavioral science and English. In behavioral science classes, subject areas such as research methods, learning, and developmental studies were considered appropriate for control groups. In English, composition, research methods, teaching methods, and linguistics were considered suitable for control groups. In experimental groups any class was potentially suitable for subjects, since only a portion of the class was selected to participate and they were aware that an experiment was in progress.

Research Design

Although the design of this study followed general aspects of the traditional paradigm, it differed with respect to the dependent measure. Whereas the dependent variable in previous studies was the amount of achievement gain, in this experiment the total number of correct items recalled became the dependent variable. Both experimental
and control groups were given the same paired-associate list to learn under different conditions. The experimental group were told they were participating in a research study, e.g., direct cue, and were moved from the usual classroom to another classroom to learn and replicate the paired-associate list. The control groups were given the same paired-associate task to learn, but in the classroom the paired-associate task became a routine assignment administered by the teacher as deemed appropriate to the curriculum. An "awareness" instrument was given to the class after completion of the paired-associate test in order to determine if subjects perceived that they were participating in an experiment, i.e., if any indirect cue was operating.

Thus, the design follows a modified post-test-only, Control Group Model, since no pre-test was necessary.

Pilot Study

A pilot study was conducted in two classes to determine (1) the appropriate level of difficulty for paired-associate items, (2) the appropriate learning and testing times to demonstrate individual differences, (3) if the instructions for the experimental groups were appropriate for the intended purpose of creating the Hawthorne Effect, and (4) if the "awareness" instrument developed for control groups communicated clearly and effectively.
The first paired-associate list used with one group of pilot study subjects was a typical list printed in an authoritative reference (1). It was abandoned when the results showed the words to be too familiar and simple in structure to be differentiating. A second, more difficult list of adjectives (see Appendix B) was designed and given to the subjects. While the adjectives were being revised, the times for learning the list were being manipulated. Since the literature on paired-associate tasks has shown concern with time needed for complete mastery of tasks, it was necessary to try different lengths of time to show optimal individual differences (1). For a ten-item list, the learning period was manipulated between three and five minutes.

A five-minute learning period was selected after a comparison of results, which indicated that the three-minute learning period was insufficient time and undifferentiating, and that five minutes was sufficient and differentiating to allow subjects to develop ways of making the associations for recall purposes, and yet to prevent frustration from occurring if the time span was too short.

Since one of the purposes of this study was to develop a climate in which the experimental group would feel they were the subjects of special attention, it was necessary to determine if these conditions were actually being met. Hence, several days after the first group of pilot subjects (N = 28)
had completed the task with the initial list, they were told that review of their performance indicated that certain members had special characteristics which would predispose them to excel on this type of task. Thus, from the original group, approximately 50 per cent were selected \((N = 12)\) to undertake another task and were moved into an adjoining room for further instructions. After the subjects were seated and told that they should do especially well at this task due to certain possessed abilities, the subjects were given the second list to learn and replicate. After the group returned to the classroom, a discussion followed, centered upon their reactions to the experiment. Immediately, the group wanted to know what were the special abilities which met the criterion of the study. The group was told that the subjects were not chosen according to any special criterion, but that their characteristics were presumably not different from those of other students in the class. The purpose for disguising the real situation was to develop conditions in which the Hawthorne Effect could occur.

Thus, the foremost concern was the selection process which supposedly differentiated the members of the group. Apparently, the experimental attempt to create an illusion of giving special treatment for certain members of the group had succeeded.

Finally, the instrument was designed to indicate if the control subjects perceived any purpose for undertaking
the task, other than being a routine class assignment. Feedback from subjects indicated that the statements communicated sufficiently well, but the response categories seemed confusing when the statement began with the first person. As a consequence, the instrument was altered to eliminate this difficulty.

A second pilot group was given the paired-associate task. The procedure differed from the first pilot group in that those conditions were similar to that of the actual study. In this situation, one class (N = 25) was used as the experimental group and the other (N = 20), the control. The control group was a developmental psychology class in which the instructor developed conditions in which the task was appropriate as a daily assignment. The other class, human development, was used as an experimental group. In the experimental situation, approximately half the class (N = 12) was used as the experimental group. The selected subjects were told that they were being asked to participate in an experiment because they had unique characteristics which would predispose them to do well in a type of learning task. The names of the selected subjects were called and they were removed to another classroom, where the task was learned and replicated by the subjects.

The results and feedback from the subjects seemed to indicate that conditions now warranted that appropriate arrangements for the actual experiment should proceed.
Faculty Orientation

As is evident from the foregoing description of the study, extensive faculty cooperation was vital to the achievement of its intended purposes. Permission was received from the chairmen of the respective departments to request assistance from individual instructors of appropriate classes in obtaining subjects. Appropriate classes for control groups were those in which the paired-associate task could legitimately fit a curriculum goal. For instance, behavioral science courses in learning, experimental methods, general and developmental psychology would have curriculum sections in which a paired-associate task was assigned as routine classroom activity. With respect to the experimental groups, any class met the stated criterion, since subjects were told they were participating in an experimental study.

In English, control classes were selected from those content areas in which a paired-associate task could be introduced as an example of how verbal material has been used to better understand the learning process in human behavior. For instance, courses in composition, linguistics, and English methods contained appropriate sections in which a paired-associate task became a class assignment. As in the behavioral science classes, English classes for experimental groups were selected without regard to curriculum appropriateness.

Orientation of participating faculty was undertaken individually. For those faculty members whose classes were
used for control subjects, extensive preparation was necessary. Since the instructor for the course in a control class was in charge of the entire testing procedure, fitting the assignment carefully into his schedule was a matter of advance planning. For the experimental classes no particular timing within the course's content was critical. In addition, the testing procedure was handled by the experimenter. (Detailed instructions for administering the test will be found in Appendices C and D.) The department chairmen agreed to suggest faculty who would be most aggreable to working on this project and to assist in arranging orientation sessions.

Selection Procedures

According to the guidelines already discussed, control subjects in both social sciences and English were complete classes in which the paired-associate task could easily be considered an appropriate activity. Experimental subjects were chosen according to a criterion as outlined later in this section.

It was determined that cells for each of the variables being examined should contain at least an N of 20. Consequently, since a sex factor underlies the examination of the other variables, a balance required a large total number of subjects. Final totals for graduate subjects were as follows: English experimental, male 22, female 22, total, 44; English,
control, male 21, female 42, total 63; behavioral science, experimental, male 22, female 22, total, 44; behavioral science control, male 27, female 22, total 49. Final totals for undergraduate subjects were as follows: English, experimental, male 23, female 21, total 44; English, control, male 26, female 30, total 56; behavioral science, experimental, male 25, female 23, total 44; behavioral science, control, male 20, female 21, total 41. The grand total of all subjects was 394.

After consulting with department heads and chairmen, it became apparent that to obtain the numbers of subjects required in each cell, subjects needed to be drawn from a number of colleges and universities in the North Texas area. As a result, the sample was taken from five area schools.

The problem of obtaining the minimum number of males and females throughout the cells was most acute at the graduate level. Graduate students were fewer in total number and there was uneven sex distribution. For instance, males predominate in behavioral science classes and females in graduate English classes.

The control classes selected were as follows: Behavioral science (graduate), research methods, human development and school psychology; behavioral science (undergraduate), learning theory; English (graduate), teaching methods, library research, and linguistics; English (undergraduate), phonetics,
composition. A total of fourteen classes and 209 subjects was used in the control group.

The experimental classes were selected on an availability basis with the cooperation of the instructor. A total of fifteen classes and 185 subjects was used.

Experimental Procedures

In control classes the instructor of each class administered the paired-associate task and test to all members of his class according to the appropriate time within his curriculum sequence. As noted, all instructors were given an individual orientation related to the purposes of the experiment and how the task could best be fitted into the class schedule. Some flexibility was necessary with respect to the manner in which the task was introduced to each class. However, it was felt that each instructor understood the importance of disguising the exercise to appear as a routine assignment to support a curriculum objective.

The experimenter conducted all sessions with experimental subjects. With the cooperation of the class instructor, each class roll was obtained for selection of subjects in advance of the class meeting. Subjects were selected to represent 50 per cent of class membership, the sex distribution varying according to minimum requirements. It was determined that no more than half the class should participate in order that those selected would feel that they were chosen according to a special criterion.
At the experimental session the complete class was told that certain members had been preselected according to certain unique characteristics to participate in an experiment, and that these unique qualities would be demonstrated on the task they were to undertake. The selected subjects were asked to leave the classroom and proceed to a nearby vacant classroom. After each group had assembled, they were given general task instructions and they proceeded with the task according to the standardized time sequence. After the testing, further questions as to the purposes of the experiment were answered. Invariably, the subjects wished to know how they were chosen. It was believed that these questions indicated they were aware that they were objects of special attention.

Analysis of Data

Data were collected from both control and experimental groups and transferred to IBM cards for computing three-way analysis of variance. The data were analyzed to permit testing the hypotheses in the null form.

A three-way analysis of variance was used because this technique permitted the examination of three independent variables, of which each contained two levels. Since no four-way analysis of variance program existed, the four independent variables of this study were manipulated in different combinations in order to answer the questions posed by the hypotheses. The four independent variables
were as follows: sex (male and female), academic level (graduate and undergraduate), discipline (behavioral science and English), and group (control and experimental).

The results of the three-way analysis of variance (1) permitted an examination of overall differences among the levels of each factor, and (2) determined whether the factors combined in such a way as to have a unique effect on the dependent variable. These two questions relate to the issues of main effects and interactions, respectively. Thus the main effects for any particular factor in a three-way analysis of variance would involve a comparison of the overall means for the levels making up that factor. Considering the four interactions associated with a three-way analysis of variance, the first three interactions show possible combinations of pairs of factors and the last interaction shows a combined relationship between all three factors.

The results of this study are reported in a series of tables related to the research hypotheses, which were either retained or rejected accordingly.
CHAPTER BIBLIOGRAPHY

CHAPTER IV

PRESENTATION OF DATA

The results of the data collected in this study will be presented by tables in relation to the hypotheses. Data were collected from students enrolled in behavioral science and English classes in five North Texas area colleges and universities. A three-way analysis of variance technique was applied in an examination of three main variables: (1) sex (male and female), (2) academic level (undergraduate and graduate), and (3) group (control and experimental). A one-way analysis of variance was applied to the data from the two academic disciplines, English and behavioral science.

Hypothesis I

The experimental subjects will achieve a significantly greater mean score on the paired-associate learning task as a measure of the Hawthorne Effect than will the control subjects. The research hypothesis was tested in the null form. When data from all eight groups were collapsed and a one-way analysis of variance was applied in a comparison of the experimental and control groups, the results as shown in Table I.
### TABLE I

**ANALYSIS OF VARIANCE FOR EXPERIMENTAL AND CONTROL GROUPS**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>12.26</td>
<td>1</td>
<td>12.26</td>
<td>1.55</td>
<td>0.21</td>
</tr>
<tr>
<td>Within</td>
<td>3100.88</td>
<td>392</td>
<td>7.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3113.14</td>
<td>393</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The means and standard deviations for the control and experimental groups are shown in Table II.

### TABLE II

**MEANS AND STANDARD DEVIATIONS FOR EXPERIMENTAL AND CONTROL GROUPS**

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>209</td>
<td>5.07</td>
<td>2.92</td>
</tr>
<tr>
<td>Experimental</td>
<td>185</td>
<td>4.71</td>
<td>2.68</td>
</tr>
<tr>
<td>Total</td>
<td>394</td>
<td>4.90</td>
<td>2.81</td>
</tr>
</tbody>
</table>

As noted in Tables I and II, the data represented do not show significant mean differences between the experimental and control groups. However, on further inspection of the three main factors involved, it was
observed that a number of interactions had developed which masked a main effect interpretation. Since the findings show that these main variables were not acting independently, this general hypothesis was not supported. Therefore, the null hypothesis was retained and the research hypothesis was rejected.

Hypothesis II

English graduate subjects will achieve a significantly greater mean score on the paired-associate learning task as a measure of the Hawthorne Effect than will English undergraduate subjects. The research hypothesis was tested in the null form. In a three-way analysis of variance for English subjects for sex, academic level and group, the data results are shown in Table III.

As observed in Table III, the main effect, academic level, was significant at the <.01 level, indicating that English graduate subjects performed significantly better than English undergraduate subjects without regard to sex or group affiliation. As stated in an earlier chapter, in formulating this hypothesis it was suggested that English graduate students would be verbally more proficient, and more cooperative with an authority figure (instructor or experimenter) in order to protect their academic image. Subjects taken from undergraduate English classes were representative of a heterogeneous group composed of students required in some degree to
TABLE III

ANALYSIS OF VARIANCE OF ENGLISH SUBJECTS FOR SEX, ACADEMIC LEVEL AND GROUP

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (A)</td>
<td>17.21</td>
<td>1</td>
<td>17.21</td>
<td>2.32</td>
<td>.13</td>
</tr>
<tr>
<td>Academic Level (B)</td>
<td>73.50</td>
<td>1</td>
<td>73.50</td>
<td>9.90</td>
<td>.01</td>
</tr>
<tr>
<td>Group (C)</td>
<td>9.43</td>
<td>1</td>
<td>9.43</td>
<td>1.26</td>
<td>.26</td>
</tr>
<tr>
<td>A x B</td>
<td>16.36</td>
<td>1</td>
<td>16.36</td>
<td>2.20</td>
<td>.14</td>
</tr>
<tr>
<td>A x C</td>
<td>4.18</td>
<td>1</td>
<td>4.18</td>
<td>0.56</td>
<td>.45</td>
</tr>
<tr>
<td>B x C</td>
<td>44.72</td>
<td>1</td>
<td>44.72</td>
<td>6.02</td>
<td>.02</td>
</tr>
<tr>
<td>A x B x C</td>
<td>.00</td>
<td>1</td>
<td>.00</td>
<td>.00</td>
<td>.99</td>
</tr>
<tr>
<td>Within</td>
<td>1477.00</td>
<td>199</td>
<td>7.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1642.31</td>
<td>206</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

take these courses and would have less vested interest in their performance. These differences, presumed to be inherent between the two English academic levels are supported by a significantly better performance by the graduate subjects.

Table IV details means and standard deviations for English graduate and undergraduate subjects.

Several interaction effects also are related to this general hypothesis. As shown in Table III an interaction existed between the variables of academic level and group identification. That is, there was an interaction between
TABLE IV
MEANS AND STANDARD DEVIATIONS FOR ENGLISH UNDERGRADUATE AND GRADUATE SUBJECTS RELATING TO ACADEMIC LEVEL

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduate</td>
<td>107</td>
<td>5.86</td>
<td>2.67</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>100</td>
<td>4.70</td>
<td>2.85</td>
</tr>
<tr>
<td>Total</td>
<td>207</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

undergraduate and graduate subjects and their identification with control and experimental groups. A further investigation involved a one-way analysis of variance between English experimental groups at graduate and undergraduate levels as shown in Table V.

TABLE V
ANALYSIS OF VARIANCE FOR ENGLISH EXPERIMENTAL GRADUATE AND UNDERGRADUATE SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>104.73</td>
<td>1</td>
<td>104.73</td>
<td>14.11</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Within</td>
<td>1477.00*</td>
<td>199</td>
<td>7.42*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Within figures taken from summary Table III.
As noted in Table V, the means square-within-figure taken from summary Table III is an estimate of the population variance based on all the data available and is the most accurate estimate for the purpose of computing significant simple effects as shown in Table V (1).

Accordingly, means and standard deviations for English graduate and undergraduate experimental groups are shown in Table VI.

### TABLE VI

**MEANS AND STANDARD DEVIATIONS FOR ENGLISH GRADUATE AND UNDERGRADUATE EXPERIMENTAL SUBJECTS**

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Graduates</td>
<td>44</td>
<td>6.59</td>
<td>2.25</td>
</tr>
<tr>
<td>Experimental Undergraduates</td>
<td>44</td>
<td>4.41</td>
<td>2.65</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In examining Tables V and VI, it is noted that English graduate experimental groups performed significantly better than English undergraduate experimental groups.

The most important finding relating to English subjects is reported in Table VII in a one-way analysis of variance between English graduate control subjects and English graduate experimental subjects.
### TABLE VII
ANALYSIS OF VARIANCE FOR ENGLISH GRADUATE CONTROL AND ENGLISH GRADUATE EXPERIMENTAL SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>39.94</td>
<td>1</td>
<td>39.94</td>
<td>5.38</td>
<td>&lt;.02</td>
</tr>
<tr>
<td>Within</td>
<td>1477.00*</td>
<td>199*</td>
<td>7.42*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Within figures taken from Table III.

Data in Table VIII reveal means and standard deviations for English graduate experimental subjects and English graduate control subjects.

The results as shown in Tables VII and VIII reveal that English graduate experimental subjects performed significantly better than English graduate control subjects. The findings lend support to the Hawthorne Effect as being operational in this specific situation. Therefore, on the basis of the results reported, the null hypothesis was rejected and the research hypothesis was accepted.

### TABLE VIII
MEANS AND STANDARD DEVIATIONS FOR ENGLISH GRADUATE CONTROL AND ENGLISH GRADUATE EXPERIMENTAL SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Graduate Control</td>
<td>63</td>
<td>5.35</td>
<td>2.83</td>
</tr>
<tr>
<td>English Graduate Experimental</td>
<td>44</td>
<td>6.59</td>
<td>2.25</td>
</tr>
<tr>
<td>Total</td>
<td>107</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis III

Behavioral science graduate subjects will achieve a significantly greater mean score on the paired-associate learning task as a measure of the Hawthorne Effect than will behavioral science undergraduates. The research hypothesis was tested in the null form. The three-way analysis of variance for behavioral science subjects is shown in Table IX. Scores of subjects taken from behavioral science classes were subjected to a three-way analysis of variance for sex, academic level, and group. Contrary to findings for subjects taken from English classes, there were no significant interaction effects, but two main effects were found significant.

TABLE IX

ANALYSIS OF VARIANCE OF BEHAVIORAL SCIENCE SUBJECTS FOR SEX, ACADEMIC LEVEL AND GROUP

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (A)</td>
<td>86.67</td>
<td>1</td>
<td>86.67</td>
<td>12.42</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Academic Level (B)</td>
<td>22.09</td>
<td>1</td>
<td>22.09</td>
<td>3.17</td>
<td>.08</td>
</tr>
<tr>
<td>Group (C)</td>
<td>49.16</td>
<td>1</td>
<td>49.16</td>
<td>7.04</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>A x B</td>
<td>5.22</td>
<td>1</td>
<td>5.22</td>
<td>.75</td>
<td>.39</td>
</tr>
<tr>
<td>A x C</td>
<td>10.28</td>
<td>1</td>
<td>.25</td>
<td>.04</td>
<td>.85</td>
</tr>
<tr>
<td>A x B x C</td>
<td>.0006</td>
<td>1</td>
<td>.0006</td>
<td>.0001</td>
<td>.99</td>
</tr>
<tr>
<td>Within</td>
<td>1249.00</td>
<td>179</td>
<td>6.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1422.63</td>
<td>186</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As noted in Table IX, main effects of sex and group are significant. Since each of these variables contains two levels, further clarification of results will be shown in Table X.
In examining Tables IX and X it is observed that without regard to academic level and group identification, females performed significantly better than did males. With respect to the control and experimental groups, the control group performed significantly better than the experimental group without regard to sex and academic level. Whereas female subjects, in general, performed better than male subjects in both behavioral science and English classes, the finding that the control groups in behavioral science performed better than the experimental groups clearly was opposite to the findings for English subjects. So, the Hawthorne Effect was not demonstrated in behavioral science.
In general, there was no support found for Hypothesis III that graduate subjects in behavioral science performed better than undergraduate subjects. Thus, on the basis of these findings, the null hypothesis was retained and the research hypothesis was rejected.

Hypothesis IV

Females will achieve a significantly greater mean score on the paired-associate task as a measure of the Hawthorne Effect than will males. The research hypothesis was tested in the null form.

As indicated in the discussion concerning Hypothesis III, behavioral science females performed better on the paired-associate task than did behavioral science males on all levels. However, among English subjects, results were more stratified. As shown in Table XI, a three-way analysis of variance was undertaken for all male subjects in relation to academic level, discipline, and group.

As indicated in Table XI, there was one significant main effect, discipline, but there were no significant interaction effects. In comparing behavioral science male subjects to English male subjects, English males performed significantly better regardless of group affiliation or academic level.
TABLE XI
ANALYSIS OF VARIANCE FOR MALE SUBJECTS BY ACADEMIC LEVEL, DISCIPLINE AND GROUP

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Level (A)</td>
<td>.96</td>
<td>1</td>
<td>.96</td>
<td>.14</td>
<td>.71</td>
</tr>
<tr>
<td>Discipline (B)</td>
<td>62.40</td>
<td>1</td>
<td>62.40</td>
<td>9.03</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Group (C)</td>
<td>2.00</td>
<td>1</td>
<td>2.00</td>
<td>.29</td>
<td>.59</td>
</tr>
<tr>
<td>A x B</td>
<td>11.56</td>
<td>1</td>
<td>11.56</td>
<td>1.67</td>
<td>.19</td>
</tr>
<tr>
<td>A x C</td>
<td>9.07</td>
<td>1</td>
<td>9.07</td>
<td>1.51</td>
<td>.25</td>
</tr>
<tr>
<td>B x C</td>
<td>5.72</td>
<td>1</td>
<td>5.72</td>
<td>.83</td>
<td>.36</td>
</tr>
<tr>
<td>A x B x C</td>
<td>12.18</td>
<td>1</td>
<td>12.18</td>
<td>1.76</td>
<td>.19</td>
</tr>
<tr>
<td>Within</td>
<td>1236.63</td>
<td>179</td>
<td>6.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1340.54</td>
<td>186</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table XII the means and standard deviation of the two groups of males are shown.

TABLE XII
MEANS AND STANDARD DEVIATIONS FOR MALE SUBJECTS IN BEHAVIORAL SCIENCE AND ENGLISH

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discipline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral Science</td>
<td>94</td>
<td>3.83</td>
<td>2.67</td>
</tr>
<tr>
<td>English</td>
<td>93</td>
<td>4.99</td>
<td>2.58</td>
</tr>
<tr>
<td>Total</td>
<td>187</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In a three-way analysis of variance, female subjects were investigated in relation to academic level, discipline, and group. Table XIII is a summary of the findings.

**TABLE XIII**

**ANALYSIS OF VARIANCE FOR FEMALE SUBJECTS BY ACADEMIC LEVEL, DISCIPLINE AND GROUP**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Level (A)</td>
<td>7.55</td>
<td>1</td>
<td>7.55</td>
<td>.97</td>
<td>.53</td>
</tr>
<tr>
<td>Discipline (B)</td>
<td>7.17</td>
<td>1</td>
<td>7.17</td>
<td>.96</td>
<td>.33</td>
</tr>
<tr>
<td>Group (C)</td>
<td>7.45</td>
<td>1</td>
<td>7.45</td>
<td>1.00</td>
<td>.32</td>
</tr>
<tr>
<td>A x B</td>
<td>98.19</td>
<td>1</td>
<td>98.19</td>
<td>13.12</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>A x C</td>
<td>9.32</td>
<td>1</td>
<td>9.32</td>
<td>1.24</td>
<td>.27</td>
</tr>
<tr>
<td>B x C</td>
<td>61.21</td>
<td>1</td>
<td>61.21</td>
<td>8.18</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>A x B x C</td>
<td>12.92</td>
<td>1</td>
<td>12.92</td>
<td>1.73</td>
<td>.19</td>
</tr>
<tr>
<td>Within</td>
<td>1489.38</td>
<td>199</td>
<td>7.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1692.88</td>
<td>206</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As noted in Table XIII, there are no significant main effects, but there are two significant interaction effects: Academic Level and Discipline, without regard to Group membership and Discipline and Group membership without regard to Academic Level.

Further analysis was required to determine the relationships between Academic Level and Discipline and Discipline
and group membership. Data in Tables XIV through XXI show the results of one-way analysis of variance for the above variables along with the related means and standard deviations. Although these tables do not relate directly to male-female comparisons as stated in Hypothesis IV, these are significant same-sex comparisons which show the extent to which the female scores were differentiated.

A one-way analysis of variance between English female subjects and academic level revealed the data in Table XIV.

TABLE XIV

ANALYSIS OF VARIANCE FOR ENGLISH GRADUATE AND UNDERGRADUATE FEMALE SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>64.75</td>
<td>1</td>
<td>64.75</td>
<td>8.65</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Within</td>
<td>1489.38*</td>
<td>199*</td>
<td>7.48*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Within figures taken from summary Table XIII.

The means and standard deviations for English graduate and undergraduate female subjects are shown in Table XV.

As observed in Tables XIV and XV, English graduate female subjects made significantly higher scores on the paired-associate task than did English female undergraduate subjects.
TABLE XV
MEANS AND STANDARD DEVIATION FOR ENGLISH GRADUATE AND UNDERGRADUATE FEMALE SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduate</td>
<td>64</td>
<td>6.22</td>
<td>2.88</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>50</td>
<td>4.70</td>
<td>2.91</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table XVI a one-way analysis of variance between academic level and discipline for females is shown.

TABLE XVI
ANALYSIS OF VARIANCE FOR ENGLISH AND BEHAVIORAL SCIENCE FEMALE SUBJECTS AT GRADUATE LEVEL

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>75.31</td>
<td>1</td>
<td>75.31</td>
<td>10.08</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Within</td>
<td>1489.38*</td>
<td>199*</td>
<td>7.48*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Within figures taken from summary Table XIII.

In Table XVII mean and standard deviation differences between graduate English female subjects and graduate behavioral science female subjects are shown.
TABLE XVII
MEANS AND STANDARD DEVIATIONS FOR ENGLISH AND BEHAVIORAL SCIENCE GRADUATE FEMALE SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discipline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral Science</td>
<td>49</td>
<td>4.57</td>
<td>2.80</td>
</tr>
<tr>
<td>English</td>
<td>64</td>
<td>6.22</td>
<td>2.88</td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From inspection of Tables XVI and XVII, it is observed that graduate English female subjects had significantly higher scores than did graduate behavioral science female subjects.

A one-way analysis of variance was applied to data relating to experimental female subjects in both English and behavioral science. The findings are reported in Table XVIII.

TABLE XVIII
ANALYSIS OF VARIANCE FOR EXPERIMENTAL ENGLISH AND BEHAVIORAL SCIENCE FEMALE SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>58.58</td>
<td>1</td>
<td>58.58</td>
<td>7.83</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Within</td>
<td>1489.38*</td>
<td>199*</td>
<td>7.48*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Within figures taken from summary Table XIII.
In relation to Table XVIII, means and standard deviations for experimental English and behavioral science female subjects are shown in Table XIX.

**TABLE XIX**

MEANS AND STANDARD DEVIATIONS FOR EXPERIMENTAL ENGLISH AND BEHAVIORAL SCIENCE FEMALE SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discipline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral Science</td>
<td>51</td>
<td>4.39</td>
<td>2.43</td>
</tr>
<tr>
<td>English</td>
<td>43</td>
<td>5.98</td>
<td>2.92</td>
</tr>
<tr>
<td>Total</td>
<td>94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As observed in Tables XVIII and XIX, female English subjects in experimental groups performed significantly better than did female behavioral science subjects in experimental groups.

In a final comparison between female subjects, a one-way analysis of variance was applied to scores of female subjects in behavioral science control and experimental groups. The results of this analysis are reported in Table XX.
TABLE XX

ANALYSIS OF VARIANCE FOR BEHAVIORAL SCIENCE CONTROL AND EXPERIMENTAL FEMALE SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>56.07</td>
<td>1</td>
<td>56.07</td>
<td>7.49</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Within</td>
<td>1489.38*</td>
<td>199*</td>
<td>7.48*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Within figures taken from summary Table XIII.

The means and standard deviations for behavioral science control and experimental female subjects are reported in Table XXI.

TABLE XXI

MEANS AND STANDARD DEVIATIONS FOR BEHAVIORAL SCIENCE CONTROL AND EXPERIMENTAL FEMALE SUBJECTS

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>42</td>
<td>5.95</td>
<td>2.82</td>
</tr>
<tr>
<td>Experimental</td>
<td>51</td>
<td>4.39</td>
<td>2.43</td>
</tr>
<tr>
<td>Total</td>
<td>93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An examination of Tables XX and XXI reveals that behavioral science control female subjects performed significantly better than did experimental female subjects.
Even though behavioral science female subjects performed significantly better than did behavioral science male subjects, a similar pattern was not found among English male and female subjects. Male English subjects were more homogeneous in their performance in relation to academic level and group than were female English subjects. As a consequence, it was not shown that English female subjects performed significantly better than did English male subjects, though female scores were considerably more heterogeneous in same-sex comparisons as noted in Tables XIV through XXI.

For behavioral science, this research hypothesis was supported. However, in English, the research hypothesis was not supported. With only partial support indicated, the null hypothesis was retained and the research hypothesis was rejected.
CHAPTER BIBLIOGRAPHY

SUMMARY, FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

The purpose of the study was to design an experiment to test the existence of the Hawthorne Effect in relation to two academic disciplines, English and Behavioral Science, using adult, graduate and undergraduate subjects and equalizing experimental and control groups with regard to sex. Since all previous Hawthorne Effect studies in education had used children for subjects and had not investigated possible sex differences in susceptibility to the Hawthorne Effect, the failure of previous research to demonstrate support for the presence of this phenomenon might be due to the imprudent willingness of experimenters to overgeneralize before there was sufficient justifying evidence. Hence, a study undertaken to examine some potentially crucial variables seemed to be justified in light of these areas of neglect in prior experimentation. As with some other social-biasing variables discussed, the Hawthorne Effect has intuitive credibility which strengthens the need for thorough examination.

Thus, the premise underlying this study was that more specific conditions should be delineated for research into
this behavioral area. As examples of an attempt to determine if students with different research experiences and subject orientation might respond differently in social experimentation, subjects were chosen from two dissimilar academic disciplines, English and behavioral science. Students in English classes were chosen as subjects because it was assumed that their limited acquaintance with social research might favor their susceptibility to feelings of being given special attention, whereas students in behavioral science classes with more experience as participants in behavioral research might be more incredulous that they had been chosen as subjects based on a "special" criterion rather than chance conditions.

Within these two environments, there might be sex and academic level differences in susceptibility to suggestion; hence, the need to equalize the sex variable with respect to male and female participation. Additionally, being at different academic levels was thought to have relevance to attitude about academic performance. It was anticipated that undergraduates would have less motivation to acquiesce to an experimenter's request for unusual productivity in which rewards were abstract, than would graduates. Therefore, both academic levels were used to test this premise.

Finally, the paired-associate task was used as an instrument for this study for several reasons. It was a verbally oriented instrument which would serve appropriately
in English control classes as an example of the bases of grammatical construction and in behavioral science control classes as an example of research on a learning task. Further, it required only a few minutes to complete, allowing the novelty to be maintained.

Findings

The results of the study appear to support the basic premise that the Hawthorne Effect may exist under specific rather than general conditions.

(1) The fact that subjects enrolled in behavioral science classes in general did not perform as well as subjects enrolled in English classes cannot be explained simply by postulating a greater verbal ability for English subjects. The instrument used did not include words which could be considered esoteric or unusually difficult for the behavioral science groups. Further, according to the Psychological Corporation profile, behavioral science students are among those with the highest verbal skills (see Appendix E). For reasons already suggested, this difference in the two groups in performance may be related directly to dissimilar attitudes about the importance of the "role" of a subject in research. If this assumption is true that subjects from behavioral science classes are less inclined to perform conscientiously in experimental situations due to over-exposure, then some attention should
be directed to this possible bias. Obviously, subjects should be chosen from a population amenable to the purposes of an experiment. Perhaps subjects should be obtained from disciplines other than behavioral science classes for behavioral research. Intuitively, it was felt that students in English classes would be more receptive and cooperative.

(2) The results from subjects taken from behavioral science classes were less stratified and were more homogeneous than were those results from subjects taken from English classes. Performances of all behavioral science groups were more similar in over-all pattern, with no significant differences in academic level.

(3) Results in behavioral science were opposite to predictions of the study in relation to demonstrating the Hawthorne Effect. As noted, the control groups performed significantly better than did the experimental groups. However, these results tend to support the speculation that, due to being more research-experienced, they were less susceptible to suggestion. Further, with respect to the better showing of the control group, it would appear that behavioral science subjects, feeling that research participation was routine and not related to their grade assessment, were not inclined to exert unusual effort, whereas for English subjects, the task exercise was more unusual and a departure from the routine.
(4) Sex differences in performance were observed in both disciplines. Female subjects from behavioral science classes performed better than did male subjects in behavioral science classes. Additionally, male subjects from English classes scored significantly better on the task than did male subjects from behavioral science classes. Female subjects from English classes showed the most heterogeneity. English female subjects in experimental groups scored significantly higher than did behavioral science female subjects in experimental groups. In addition, female subjects from English graduate classes scored significantly better than subjects from English undergraduate classes. Finally, English female graduate subjects assigned to experimental groups scored significantly better than behavioral science female graduate subjects assigned to experimental groups. It seems apparent that specific rather than general conditions persist. Female graduate students enrolled in English classes and assigned to experimental groups appeared to be the most proficient specific group which tended to support the Hawthorne Effect phenomenon.

(5) The most important result for the purposes of this study was the significantly better performance of the English graduate experimental group over that of the English graduate control group. This would be evidence supporting the Hawthorne Effect. Since no previous research in education has demonstrated this phenomenon, this finding has
important implications. In light of the 1976 article by Horne (2), suggesting the unreported motivation for monetary gain as the real condition present in the original research, the results of the present study are unusually provocative.

Conclusions

The findings of this study on the Hawthorne Effect have tended to support the notion of the specific rather than the general nature of the Hawthorne Effect. Differences in performance favored the notion that a non-research-oriented adult graduate female would be the most likely subject to demonstrate this condition. Results are less clear with respect to the expected performance of male subjects enrolled in English graduate classes. As speculated, since behavioral science and English as disciplines showed opposite results under the same research conditions, attitudes toward research may account for some of the disparity. Out of an array of possible adult subjects on a college campus, perhaps subjects taken from disciplines other than behavioral science might prove more productive.

Recommendations for Further Research

The findings of this study tend to keep alive the notion that the awareness of being considered unique in some manner does motivate performance beyond the normal base level under specific conditions and with specific types of subjects.
In light of the present research, the following recommendations are made:

1. The personality variables of the researcher and those of the subjects being tested should be delineated with greater precision to discover what combinations in concert may produce the Hawthorne Effect. Braun presented a personality model relating to the teacher-expectancy concept which may be appropriately redefined for use in the Hawthorne Effect research (1).

2. Although the results of this study were inconclusive regarding sex differences in susceptibility to the Hawthorne Effect, there is considerable indication that certain types of females may be affected. The finding that females in English were highly different in performance was dramatic in that behavioral science females did not show a stratified tendency. However, behavioral science females did perform significantly better than did males, whereas English female subjects were not clearly superior to English male subjects. Consequently, research to demonstrate possible sex differences in susceptibility to the Hawthorne Effect should be continued.

3. Since contrasting results were obtained from subjects in English and behavioral science academic areas, future studies using academic settings should focus upon other disciplines, preferably those in which students are
unaccustomed to participating in research studies. The present study suggested that research-naive subjects were more susceptible to the Hawthorne Effect than were research-experienced subjects.

4. Other social settings should be examined for possible contamination of research results due to the Hawthorne Effect. For example, since the results of the original study appear to be questionable, further industrial research would be warranted.

APPENDIX A

VERBAL LEARNING TASK AND
VERBAL LEARNING TEST FORM
VERBAL LEARNING TASK

DIRECTIONS: The following is a list of ten adjectives paired with a trigram. You will be given five minutes to learn these terms. At the end of this time period you will be asked to replicate this material from memory.

1) VOM - Recalcitrant
2) NAC - Flaccid
3) TAH - Ambitious
4) VEZ - Deceitful
5) TEH - Sarcastic
6) WEP - Awkward
7) NTC - Exuberant
8) STJ - Memorable
9) WAP - Tumultuous
10) SUJ - Fearful
TEST FORM

VERBAL LEARNING TASK

DIRECTIONS: To the following trigrams attach as many of the correct adjectives as you can. You will be allowed one minute.

1) TEH
2) SUJ
3) VOM
4) NAC
5) VEZ
6) NTC
7) TAH
8) WEP
9) STJ
10) WAP
APPENDIX B

STUDENT REACTION INSTRUMENT
STUDENT REACTION INSTRUMENT

NAME______________________________________

DIRECTIONS: In the space to the left of each statement place the letter representing most closely your feelings regarding your participation in this assignment.

(1) It was anticipated that a course of this type would have assignments of this nature.
(2) It is important to study the learning process in order to better understand myself.
(3) From the nature of the verbal learning assignment it seemed clear that this task was a good learning experience.
(4) The teacher treated this assignment no differently than he/she did others during this course.
(5) It was not anticipated that a course of this type would have assignments of this nature.
(6) Even though the teacher said this assignment did not affect my grade, I felt that a poor performance would influence the teacher's opinion of me.
(7) The teacher treated this assignment differently than she did others during the course.
(8) From the nature of the verbal learning assignment, it seemed clear that the class was participating in an experiment.
(9) It is not important to understand the processes involved in verbal learning.
(10) This type of learning experience was enjoyable because it was a change of pace from the usual classroom activity.
RESPONSE CATEGORIES

A. I know this is true.
B. I suspect this is true.
C. I cannot decide.
D. I suspect this is not true.
E. I know this is not true.
APPENDIX C

INSTRUCTIONS GIVEN TO CONTROL GROUP
INSTRUCTIONS FOR CONTROL GROUP

In college we are expected to learn new material which often seems unrelated to information we have already learned. As English (psychology) scholars we are interested in a better understanding of the way we go about learning new and unrelated material. In order to demonstrate this process, I'm going to ask each of you to participate in a special kind of verbal learning task. In order that all of us start from an initial learning base, I am going to substitute nonsense syllables for other kinds of verbal materials. I will pass out this material and you will be allowed five minutes to learn the lists of words. Afterward you will be given a form on which you will replicate the verbal material you have learned.

Even though this will not affect your final grade, please try to do the best job that you possibly can.

Delay any questions until the assignment is completed.

NOTE TO INSTRUCTOR: Material should be given to students face down until all have received copies and signal is given to turn the copy over as timing begins. After learning period is completed, take up copies and give students test copy face down until signal for timing begins. Allow one minute on testing. Then take up test forms and place all material in brown envelope.
INSTRUCTIONS FOR EXPERIMENTAL GROUP

I am going to ask a number of you to participate in a very important experiment involving a verbal learning task. This is an important experience because it is anticipated that the results will be instrumental in changing the nature of future experimental design. We have pre-selected a group of you to participate because we are looking for certain personal characteristics which relate to optimal learning in this type of task. Those of you who are selected have these unique characteristics.

I am going to ask that those of you whose names I call move into room ___ where I will give you further instructions. (These instructions to be given after experimental group is seated).

I am going to pass out this material and you will be allowed five minutes to learn list. Afterward you will be given a form on which you will replicate the verbal material you have learned.
APPENDIX E

THE PSYCHOLOGICAL CORPORATION PROFILE FOR MAT
Your test raw score can be compared with the following abbreviated norms. This comparison is limited in meaning for several reasons: Different schools, agencies and employers set their own standards, and these standards vary widely. A score that is high for one university may be only average for another. The weight placed on test scores in the making of decisions varies widely among institutions and situations. Neither the examiner nor the publisher can take further responsibility for interpreting scores or advising individuals. Consult your own adviser, counseling bureau, or personnel department.

Raw Score Equivalents of Three Percentile Points
(Applicable to all forms of each test except where specified)

<table>
<thead>
<tr>
<th>Test</th>
<th>Group</th>
<th>Major Field</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>25th</td>
<td>50th</td>
</tr>
<tr>
<td>MAT</td>
<td>First-year graduate students</td>
<td>Bus. Admin.</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Education (D)</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Education (M)</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Educ. Admin.</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Engineering</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nursing</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Psychology</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social Work</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theology</td>
<td>33</td>
</tr>
<tr>
<td>MEAT</td>
<td>Applicants to Graduate School (Form E)</td>
<td>Engineering</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Applicants for Engineering Positions (Form F)</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Employed Engineers (Form E)</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>DMR</td>
<td>Senior Students</td>
<td>Psychology</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Graduate Students</td>
<td>Psychology</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Education</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Applied and Math.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Statistics</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Employed Engineers</td>
<td></td>
<td>34</td>
</tr>
</tbody>
</table>

(D) Institutions granting doctoral degrees,
(M) Institutions granting master's only.

The table is read as follows: Among first-year graduate students in Business Administration, 25 per cent of the group obtained a Miller score of 43 or less; the median (middle) score of the group was 55; and those with scores above 65 were in the upper quarter.
BIBLIOGRAPHY

Books


Wise, Jere D., Test Anxiety and Self-Attention, Toronto, Ontario Mental Health Foundation Press, 1974.

Articles


Linder, Steve and Carol Whitehurst, "Is There a Novelty Effect on Student Attitudes Toward Personalized Instruction?" The Journal of Experimental Education, XLII (Fall, 1973), 42-44.


Meyer, William J. and George G. Thompson, "Sex Differences in the Distribution of Teacher Approval and Disapproval Among Sixth-Grade Children," Journal of Educational Psychology, XLVII (1956), 385-396.


Pennock, Clifford D., "Student Teacher Expectations for Primary Level Boys' Reading Achievement," Dissertation Abstracts International, XXXII (1962), 5657-A.


Reports


The Relationship Between Possible Hawthorne Effect Components and Stages of Experimental Investigation, Columbus, Ohio, Ohio State University, Bureau of Educational Research Services, 1963.

Dalis, Gus T., Using Precise Objectives to Enhance Student Achievement in Health Education, report to Los Angeles County Superintendent of Schools, Los Angeles, California, 1970.


Publications of Learned Organizations


Klausmeier, Herbert J. and Katherine V. Feldman, The Effects of a Definition and a Varying Number of Examples and Non-Examples on Concept Attainment, Research and Development Center for Cognitive Learning, University of Wisconsin, Madison, Wisconsin, October, 1973.

Unpublished Materials


Peterson, John C., "Effect of Exploratory Homework Exercises Upon Achievement in Eighth Grade Mathematics," unpublished doctoral dissertation, Department of Education, Ohio State University, Columbus, Ohio, 1969.
