5-1968

Differential Hawthorne Effect by Cueing, Sex, and Relevance

Richard Carl Harris
Utah State University

DIFFERENTIAL HAWTHORNE EFFECT BY CUEING, SEX, AND RELEVANCE

by

Richard Carl Harris, Jr.

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Psychology

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1968

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

ACKNOWLEDGMENTS

Sincere appreciation is extended to the writer's graduate committee: Dr. David R. Stone, Dr. Keith Checketts, and Dr. Ross Allen all of whom contributed both time and constructive ideas to this research project. Dr. James P. Shaver was most helpful with his suggestions on design and analysis.

Appreciation is also extended to Richard Sweetland and Helen Mitchell whose cooperation in allowing the experimenter to work with their students was absolutely essential to the completion of the study.

This writer also extends deepest gratitude to his wife, Melanie, whose efficient typing and proofreading made the completion of this thesis possible.

Richard Carl Harris, Jr.

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Differential Hawthorne Effect by
Cueing, Sex, and Relevance
by
Richard Carl Harris, Jr., Master of Science
Utah State University, 1968

Major Professor: Dr. David R. Stone
Department: Psychology

This study attempted to create experimentally the Hawthorne effect in a freshman general psychology class at Utah State University during fall quarter of 1967. It also attempted to discover the differential effect of cueing, sex, and relevance on the experimental creation of the Hawthorne effect as measured by six general psychology criterion tests.

The design of this study included a control group and three experimental groups.

The following five hypotheses were postulated:

1. The experimental groups will show greater influence from the Hawthorne effect than the control group.

2. Within the three experimental groups there will be an increasing Hawthorne effect with the least effect in the subject-object cue group and the greatest effect in the subject-object-observer group as compared to the control group.

3. The females in all experimental groups will show significantly greater Hawthorne effect than males within the same groups.

4. The group rating high on the Relevance scale will show significantly greater Hawthorne effect than the groups rating low in relevance.

5. There will be sufficient interactive effects between factors to the extent that some will reach significance.

The hypotheses were tested by means of analysis of covariance with ACT predicted grade point average as the covariate. None of the differences were significant at the .05 level.

It was concluded that the Hawthorne effect does not exist as a potent enough variable to distort the influence of the independent variable on the dependent variable in educational and psychological investigations of short duration involving freshman university students. It was also concluded that the variables of Cueing, Sex, and Relevance are not functionally related to the creation of the Hawthorne effect and, therefore, need not be controlled.

(52 pages)

INTRODUCTION

Background of the Problem

In classrooms, industrial plants, therapeutical clinics, and playrooms, educators and psychologists are dealing with learning. Their concentration is usually directed towards expanding and improving present knowledge and techniques about what learning is and how to facilitate it.

After educators and psychologists develop new methods in an attempt to facilitate learning, they put their new methods in the classroom to test their effectiveness. Ordinarily they will try out new methods on an experimental group and compare this group's learning gains to that of a control group which has been exposed to the ordinary method.

In many studies using this design, a phenomenon may take place much as Harold F. Clark explains in regards to experiments involving phonics vs. word meaning for improving reading.

> Experiments that have emphasized phonics have brought improvements in reading; experiments that have emphasized the meaning of words but with less emphasis on phonics have also brought increases in reading ability. Comparable experiments have been conducted in all major subject matter fields. Again, the most reasonable interpretation is that there are strong experimental interest-enthusiasm factors at work. (Clark, 1963, p. 48)

As Clark implies, educators and psychologists may compare their new methods of learning reading skills, but not know whether to attribute the improvement to phonics, word meanings, or "strong experimental interest-enthusiasm factors."

Another researcher, Cronbach, alludes to the same problem encountered in curriculum evaluation in order to bring about needed course improvement.

> In an educational experiment, it is difficult to keep pupils unaware they are an experimental group. And it is quite impossible to neutralize the biases of the teacher as those of the doctor are neutralized in the double-blind design. It is thus never certain whether any observed advantage is attributable to the educational innovation as such or the greater energy that teachers and students put forth when a method is fresh and experimental. (Cronbach, 1964, p. 237)

This effect which obscures the influence of the independent variable in a dependent-independent variable design study has been variously referred to as the novelty effect, awareness, attention, interaction, motivation, Hawthorne effect, and similar synonyms (Cook, 1967, p. 8-10).

The effect, which will be referred to as the Hawthorne effect (see E. Mayo's, The Human Problems of an Industrial Civilization, for a history) was defined by Desmond L. Cook as follows:

> The Hawthorne effect is a phenomenon characterized by an awareness on the part of the subjects of special treatment created by artificial experimental conditions. This awareness becomes confounded with the independent variable under study, with a subsequent facilitating effect on the dependent

> variable, thus leading to ambiguous results. (Cook, 1962, p. 118)

Cook used this definition of the phenomenon on which to base a three-year study on the impact of the Hawthorne effect in experimental designs in educational research. His study set out intentionally to create the Hawthorne effect using new curriculum materials developed by the School Mathematics Study Group (SMSG) as compared to regular materials. Cook's general finding was that no significant differences were produced between the gains of the experimental groups and control groups. No significant correlations were found between awareness of experimental participation and achievement (Cook, 1967, p. 100). The implications of Cook's study are summarized in this paragraph:

> One distinct possibility is that the variable under concern, the Hawthorne effect, simply does not exist as a variable of sufficient potency to be significantly influential on study results. On the other hand, it is possible that the phenomenon may exist but that it spreads equally over all treatment conditions, and thus its influence in an experimental investigation is minimized if not altogether eliminated. If either of these interpretations possess validity, educational researchers could proceed to disregard the possible operation of the phenomenon in their investigations and accept significant differences as being basically due to the independent variable (s) introduced as part of the experimental design. (Cook, 1967, p. 100)

Statement of the Problem

The problem then is our lack of information about the Hawthorne effect and its influence on learning. The justification that such lack of information is a problem comes in Cook's concluding remarks on his three-year study.

> If educational researchers sincerely believe, and the principal investigator believes they do, that the Hawthorne effect is a variable causing them great concern in the conduct of their research, it would seem any time and energy be devoted to securing valid and reliable evidence with regard to the concept would be justifiable. (Cook, 1967, p. 124)

This general problem of lack of information can be broken down into several specific problems or question areas.

1. Will the Hawthorne effect be potent enough after three months to distort the influence of the independent variable? (Cook, 1967, p. 101, 105, 106)

2. Will cues have differential effects in creating the Hawthorne effect? (Cook, 1967, p. 117; Orne, 1962, p. 779; Riechen, 1962, p. 31; Dixon, 1966, p. 155)

3. Will sex play a significant role in the creation of the Hawthorne effect? (Cook, 1967, p. 104; Roethlisberger, 1940, p. 20; Gall and Mendelsohn, 1967, p. 216; Tyler, 1965, p. 259)

4. Will the degree to which the subjects perceive the experimental situation to be relevant in terms of scientific research directly influence the Hawthorne effect? (Orne, 1962, p. 777; Frank, 1944)

5. Will interaction between experimental variables significantly influence the Hawthorne effect?

Purpose of the Study

The purpose of this study was to discover if the Hawthorne effect could be experimentally created in a freshman general psychology class as measured by six multiple-choice criterion tests. If the effect was created, then it was the purpose of this study to discover the influence of subject-object, subject-observer, and subject-object-observer cues; the influence of sex; the influence of perceived relevance; and the influence of interaction between the experimental variables on the creation of the Hawthorne effect.

Definition of Terms

Independent Variable. The teaching methodology used in the general psychology course.

Dependent Variable. The student achievement as was measured by six multiple-choice psychology tests.

Cues. This referred to subject-object cueing, subject-observer cueing, and subject-object-observer cueing.

Subject-Object Cue. This cue was presented to the subjects or students in the form of a wall mural, teaching machine, lights, record player, camera, colored paper, and lettered paper.

Subject-Observer Cue. This cue was the investigator sitting throughout the class period observing and taking notes from a clearly obvious vantage point.

Subject-Object-Observer Cue. This cue was a combination of both the other cues. For example, both the lights and observer acted as cues for the same group of subjects.

Relevance. This meant the importance that was attributed by the subjects to the experimental situation in terms of contributing or playing a significant role in advancing scientific research.

REVIEW OF THE LITERATURE

Literature Related to Background

The phenomenon known as the Hawthorne effect is by no means new in educational research. As early as 1923 McCall was referring to it in these terms:

> Though evidence on this question is meager, there is some reason to believe that the mere process of experimenting with new methods or materials of instruction attracts such attention to the traits in question as to cause an unconscious concentration both on the part of teacher and pupil upon progress in these traits. (McCall, 1923, p. 67)

McCall's statement shows that the Hawthorne effect phenomenon was recognized in educational research possibly before it was recognized in industrial research.

Because the effect's nomenclature originated in industrial research it is appropriate to give a brief accounting of those studies. In 1924 the Massachusetts Institute of Technology initiated a series of tests under the sponsorship of the National Research Council and the Illuminating Engineering Society to ascertain the relationship between illumination and production in various factory situations (Snow, 1927). After the initial investigations in all the different plants no direct relationship was found between illumination and production. Snow gives some insight as to why.

> Many of them (contaminating variables) can be controlled or eliminated, but the one great stumbling block remaining is the problem of the psychology of the human individual. (Snow, 1927, p. 282)

Although there were many illumination experiments, the Hawthorne investigations became better known because at this point, the Hawthorne researchers undertook a series of investigations designed to develop ways of studying the introduction of variables into work situations. Consequently, rest periods, working hour changes, and wage incentives were introduced and observations made of the worker's reactions to such variables. Pennock reports on the observations.

> From these tests have come startling results; startling because they were unexpected as well as because they were sometimes contrary to accepted opinion. In thefirst place, there was a gradual yet steady increase in production regardless, to a certain extent, of test conditions imposed. (Pennock, 1929, p. 304)

Pennock states that although several hypotheses were suggested, all were rejected for lack of significant relationships. What he considered to be the major accomplishment of the entire study was this:

> ...which leaves us convinced that the rather remarkable results we have been able to obtain with this group are due mainly to changes in their (female workers) mental attitude. (Pennock, 1929, p. 309)

The impact of this finding was so marked that it led the Western Electric Company to explore the nature of employee attitudes. This latter group of studies was to become highly influential in the development of the area of industrial psychology now known as "human relations in industry." A complete account of the various investigations,

including the illumination experiments of Hawthorne, is presented in the book, Management and the Worker by Roethlisberger and Dickson (1941). A retrospective view of the investigation and subsequent implications for industry are presented in Hawthorne Revisited by Landesberger (1958).

The preceding is relevant in that it gives a basis for defining the effect under study as well as pointing out the similarity between illumination investigations and contemporary educational research. Cook points out this parallel.

> A change is introduced and promising results are secured. This promising lead is followed up by carefully controlled experimentation to study more precisely the effects of the change. The results are too often similar to those obtained in the illumination experiments. Regardless of what is done, we have difficulty in attributing observed changes in the dependent variable directly to the manipulated independent variable. (Cook, 1962, p. 118)

Based on this thinking as well as a partial definition put forth by French (1953, p. 101), Cook gives a working definition quoted in the "Background of Problem" section of this thesis.

Literature Related to Problems

This review of relevant literature has to this point briefly outlined the conceptualization of the Hawthorne effect and laid a basis for a definition. From this vantage point this review will look to Cook's report for the origins of the problems mentioned in the "Statement of Problem" section. It will also relate to each problem other significant thinking as viewed by other writers.

The first problem deals with the influence of time on the potency of the Hawthorne effect. In this regard Cook points out:

> The findings presented are, however, limited because no measurements were taken within the first month or two to see if there were early significant differences in achievement between treatment combinations which were then reduced or eliminated at the end of one or two years. (Cook, 1967, p. 106)

Not only does the question remain unanswered in regard to Cook's study but contradictory evidence seems to exist as Cook (1967) points out from the literature analysis portion of his study.

> Data from the literature analysis with regard to the duration of the study and experimental results revealed a relationship contrary to what seems to be a logical position that novelty and similar effects would diminish over time. The evidence accumulated here indicated that the longer the study was conducted the more significant differences favoring experimental over control groups as contrasted to equality between the two groups or the control exceeding experimental groups were observed. (Cook, 1967, p. 116)

The second problem, concerning type of cueing, finds root in Cook's statement.

> It appears unlikely that one can employ a Hawthorne effect concept to explain differences or the lack of differences between experimental and control groups in educational research studies in so far as the variable commonly believed to generate the effect such as direct and indirect cues, the duration of the study, and mechanical changes introduced in an experiment are considered to be of sufficient potency to produce the effect. (Cook, 1967, p. 117)

Though Cook feels cues have little effect, Orne (1962 b) feels that the totality of cues which convey an experimental hypothesis to the subject become significant determinants of the subject's behavior. They have labeled the sum total of such cues as the "demand characteristics of the experimental situation." Riechen (1962) also feels that cueing has significant impact on subject behavior.

> Now, the subject is not such a fool as the experimenter wants to make him out. He suspects that various answers are right and wrong to the extent that they represent him to the experimenter in the light that he (the subject) wishes to appear--that there are answers that will enhance and that will diminish his values as a person. (Riechen, 1962, p. 31)

Another researcher, Dixon (1966) points out from his study on experimenter-subject relationship that where there existed cueing connoting an "impersonal" relationship no appreciable conditioning resulted whereas when cueing denoted a personal experimenter-subject relationship, marked conditioning resulted.

Cook feels that cues resulting from the continuous or frequent presence of external persons such as the principal investigator would have a negative influence on the experiment by perhaps highlighting the situation. Cook, as the principal investigator, purposely remained obscure until the very end of the study when he interviewed the participating teachers. Even in light of this, he still indicates the need for more information about the influence of such a cue in this statement:

> The question of how much and in what manner the presence or absence of the principal investigator in an experimental situation such as the one conducted in this project might well be the subject of further study. In the present study perhaps he should have appeared more often in selected classrooms in order to heighten the desired effect.(Cook, 1967, p. 37)

The second problem then becomes one of shedding more light on these differing positions concerning the influences of "object" cueing as compared to "observer" cueing.

The third and fourth problems are suggested by Cook in his statement made in trying to explain why the Hawthorne effect did not influence the independent variable in a measurable manner.

> A further possible explanation of the results might be that the Hawthorne effect does exist within students and classes, but it might well be considered as a variable of insufficient potency to cause any real differences. This might possibly explain why the relationships between awareness and gains varied between classroom groups within treatments. That is, within a given single treatment combination one classroom group might have a positive relationship between gain and awareness while the second classroom might have a negative relationship.(Cook, 1962, p. 104)

It seems from Cook's statement that what is needed is information about the dynamics operating within treatment groups as well as between groups.

One dynamic (which becomes problem three) apparently not considered by Cook is the sex of the subject involved. It is noteworthy that in the original Hawthorne studies females were used as subjects as is indicated in this statement by Roethlisberger.

> The job finally chosen as best fulfilling these requirements was the assembly of telephone relays, an operation performed by women.(Roethlisberger, 1940, p. 20)

Pennock also alludes to the "girls" in talking about the study. That the Hawthorne effect may be considerably more potent with female subjects than males, finds basis from many researchers.

> The greater effect of the social-psychological aspects of the experiment on female subjects than on males is to be expected in light of the extremely consistent findings that females are more dependent upon, sensitive to and responsive to other people than males. (Gall and Mendelsohn, 1967, p. 216)

Tyler (1965, p. 259) suggests, also, the increased "sensitivity-responsiveness" of females to environmental changes. Other researchers such as Oetzel (1962), Goodenough (1955), Patel and Gordon (1961) and Bennet and Cohen (1959) support this general concept.

Problem four, also concerned with the underlying dynamics of the experimental situation, deals with the degree to which the adult subject sees the experiment to be relevant and important to the advancement of science and perhaps ultimately to human welfare in general (Orne, 1962).

In pilot studies conducted by Thomas Menaker but reported by Orne, subjects were given clearly impossible tasks to complete. After the instructions were given, the subject was deprived of his watch and told, "Continue to work; I will return eventually." Not until five and one-half hours later did the subject give up. Even on tasks

the experimenters thought would be discontinued spontaneously within a brief period, subjects persisted with little sign of overt hostility. When the experimenters gave post-experimental interviews, an explanation was furnished.

> When asked about the tasks, subjects would invariably attribute considerable meaning to their performance, viewing it as an endurance test or the like. (Orne, 1962, p. 777)

This observation is consistent with Frank's (1944) failure to obtain resistence to disagreeable or nonsensical tasks. That this dynamic of the experimental situation is a problem and worthy of study is supported by Orne.

> The study and control of demand characteristics are not simply matters of good experimental technique; what circumstances demand characteristics significantly affect subjects experimental behavior. (Orne, 1962, p. 783)

Problem four is one of knowing whether more relevance or significance is attributed to the experimental situation where object cues are given as compared to observer cues.

Problem five was concerned with interaction between variables. Interaction is defined by Campbell as, "the law as to the effect of A changes depending upon the specific value of B" (Campbell, 1963, p. 199). Shaver (1967) has indicated that in the past many potentially significant findings have been omitted because of the failure on the part of experimenters in educational research to pay little or no attention to interactive effects between variables. In this study interaction between Cueing, Sex, and Relevance was examined for pertinent dynamics.

PROCEDURE

Hypotheses

1. The experimental groups will show greater influence from the Hawthorne effect than the control group.

2. Within the three experimental groups there will be an increasing Hawthorne effect with the least effect in the subject-object cue group and the greatest effect in the subject-object-observer group as compared to the control group.

3. The females in all experimental groups will show significantly greater Hawthorne effect than males within the same groups.

4. The group rating high on the relevance scale will show significantly greater Hawthorne effect than the groups rating low in relevance.

5. There will be sufficient interactive effects between factors to the extent that some will reach significance.

Sample

The sample consisted of 134 students who had enrolled in the general psychology course offered by the Psychology Department at Utah State University for fall quarter, 1967. In accordance with departmental regulations, the students signed up for one of four laboratory periods[1] each of which met once a week for 50 minutes. Neither random selection nor assignment was possible in selecting subjects for the groups in that the experimenter was permitted no controll over registration procedures.

Design

This factorial design was constructed with reference to Campbell's (1963) statements on experimental and quasi-experimental design for research on teaching. It resembles what Campbell calls a "post test only, control group" design. Each week for nine weeks, three of these four groups of subjects were exposed to different cues all of which were intended to create the aura of experimental conditions and special treatment. The treatments were placeboes in that they had no bearing or relationship to the ongoing class procedures (Fillenbaum, 1966; Travers, 1958).

---

[1]The words "laboratory period" should not be construed to mean a period of experiments after the manner of the natural sciences. The only things that differentiated these periods from regular lecture sessions was that the groups were smaller and the teaching assistant took charge.

Cueing in the groups was as follows: Cue-N group received regular classroom activities. Cue-1 group received regular classroom activities plus subject-object cues. The subject-object cues followed somewhat the various environmental changes that were performed in the original Hawthorne studies (Mayo, 1946; Roethlisberger, 1947). Cue-2 group received regular classroom activities plus the direct observation of the experimenter, he being the subject-observer cue. Cue-3 group received regular classroom activities plus the combination of the subject-object cues and the subject-observer cue. See Table 1.

All three treatment groups were informed the first day of class that they were part of a study, but no further mention was made about the study until the last day of class.

All groups had the same instructor and the same materials in class. All groups were tested six different times during the quarter on psychological concepts as learned in the course (see Instrumentation section and Table 2). The last day of the course all groups were given a questionnaire. This questionnaire gathered pertinent biographical data and attempted to assess what importance the subjects had attributed to the experimental situation (see Instrumentation section and Appendix).

This design was constructed to deal with three factors: Cue, Sex, and Relevance. The Relevance factor was meaningless for those subjects in the control group as they had

Table 1. Treatment schedule of cues.

Date	8:30 a.m. Cue-1 Subject-Object	9:30 a.m. Cue-2 Subject-Observer	10:30 Cue-N No cue	1:30 p.m. Cue-3 Subject-Object-Observer
Sept 28	Lights	Observer	None	Lights and observer
Oct 5	Green paper	Observer	None	Green paper observer
Oct 12	Lights	Observer	None	Lights and observer
Oct 19	Green paper with printing	Observer	None	Green paper with printing and observer
Oct 26	Camera	Observer	None	Camera and observer
Nov 2	Record player	Observer	None	Record player and observer
Nov 9	Reading machine	Observer	None	Reading machine and observer
Nov 16	Wall mural	Observer	None	Wall mural observer
Dec 7	Questionnaire	Questionnaire	Questionnaire	Questionnaire

received no indication or experimental activities. If relevance existed, it could have no connection with this study; therefore the Cue-N level is omitted.

To handle this irregularity sub-designs A and B were employed (see Figures 1 and 2). Sub-design A took into account the Cue and Sex factors. On the Cue factor, four cue levels (Cue-N, Cue-1, Cue-2, and Cue-3) were considered. On the Sex factor two levels (Male and Female) were considered.

Sub-design B dealt with three factors: Cue, Sex, and Relevance. The Cue factor had three levels (Cue-1, Cue-2, and Cue-3). The Sex factor had two levels (Male and Female). The Relevance factor had two levels (High and Low).

Sub-design A was used to seek information concerning problems one and two, and therefore hypotheses one and two. It was hoped that in this part of the design the question as to whether a Hawthorne effect could be created with increasingly more potent cueing could be answered. Sub-design B was meant to seek information dealing with the effect of "investigator presence," "female susceptibility," and "demand characteristics" of the experimental situation." It was hoped that both sub-designs could give insight into interactive effects.

Sub-Design A

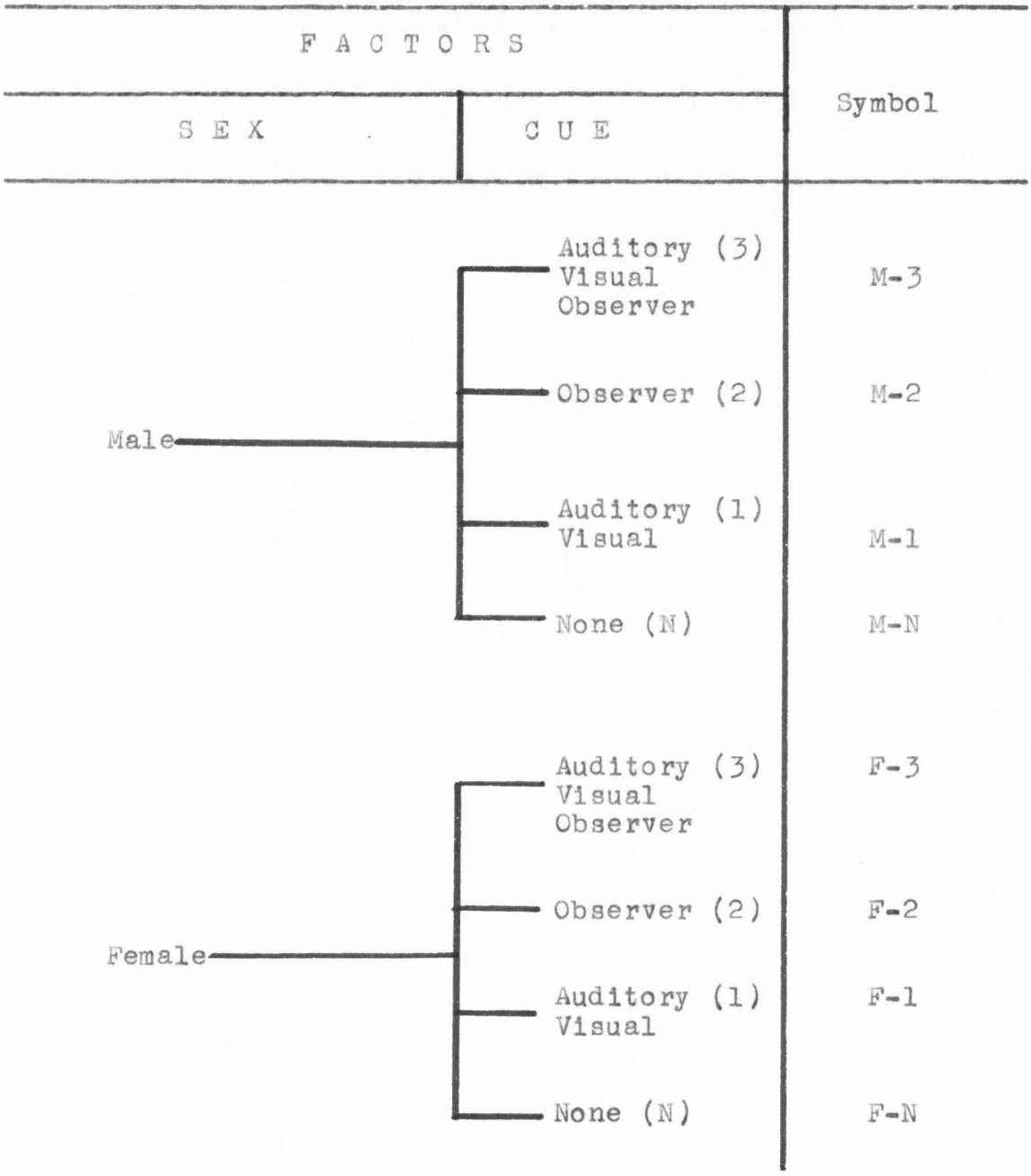

Figure 1. Experimental design variables and treatment combinations.

Sub-Design B

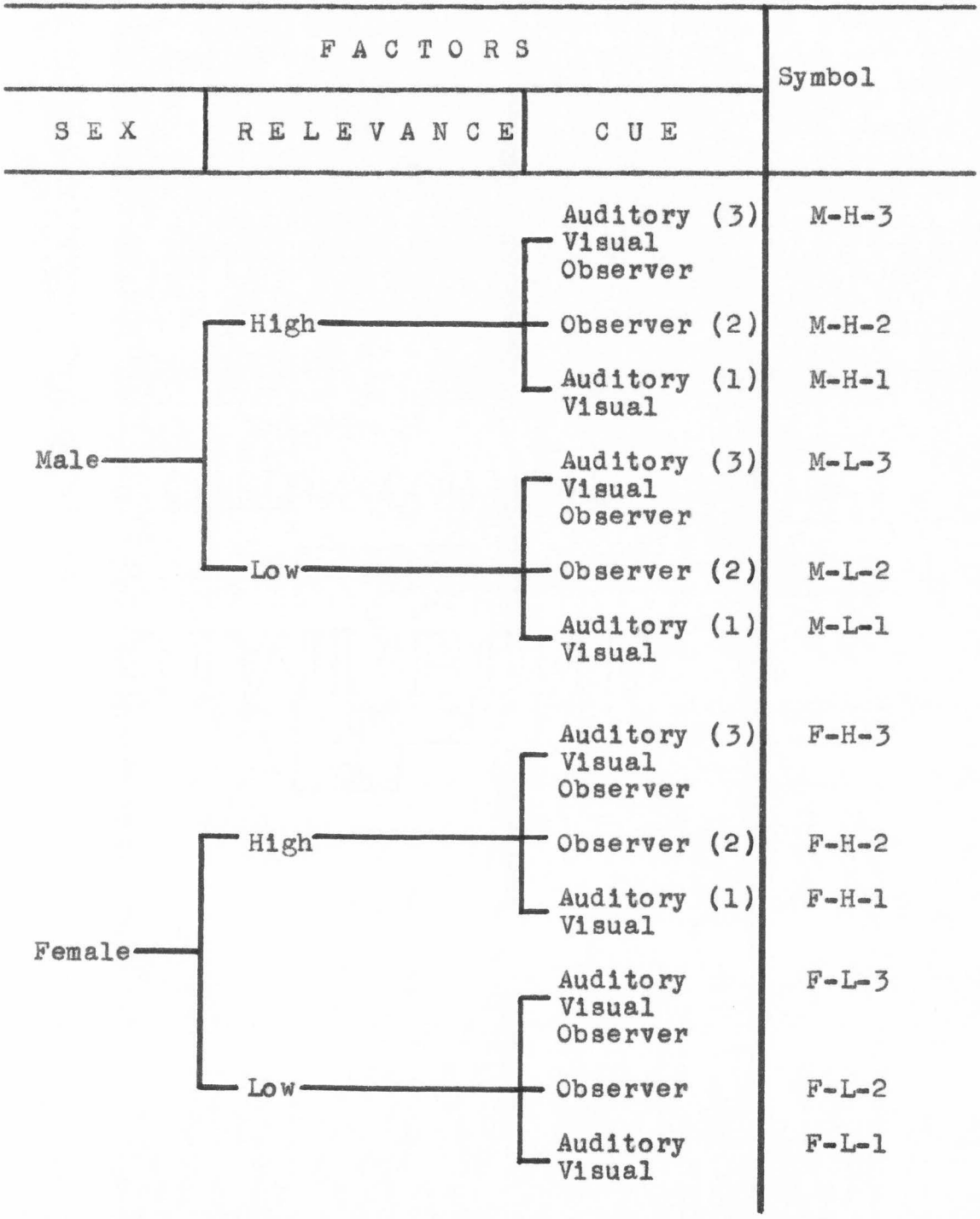

Figure 2. Experimental design variables and treatment combinations.

Instrumentation

Six achievement tests of the objective, multiple-choice type were administered to each group in the third, fourth, sixth, eighth, and eleventh weeks of the quarter. These tests were based on the textbook, Psychology and Life, by Ruch, Warren, and Gorfein (1967). The first five tests dealt with specific chapters of the text while the sixth test was comprehensive through the complete book and course. Each test was constructed by the instructor of the course from the teacher's manual which accompanied the text. The laboratory instructor, who met with the four groups once a week, had no knowledge of the test questions until after the administration of each test. This guarded against any teaching bias based on knowledge of test questions.

To measure the perceived relevance, part of the questionnaire that each subject filled out the last day of class contained a five-point scale. This perception scale ranged from, "The experiment appears worthless" at one end to "The experiment appears very valuable" at the other (see Appendix). Three categories were tabulated; the two at either extreme and a middle category. When tabulation was attempted for these three categories, it was found that there was a very small number in the lowest category. It was decided at this point to collapse the low relevance category and use the middle and high categories for measuring low and high relevance. Those few who had rated in the original low category were transferred to the middle category.

Table 2. Instrumentation and data collection schedule.

Date	Instrument administered
October 10	14 question multiple choice test
October 12	60 question multiple choice test
October 20	60 question multiple choice test
November 14	60 question multiple choice test
December 6	60 question multiple choice test
December 7	Questionnaire
December 11	120 question multiple choice test

To reduce the effect of initial group differences, demonstrated academic ability was held constant as a covariate. This covariate was measured by the American College Test (ACT) which is administered to all entering freshmen by the Utah State University Testing Services. The experimenter obtained from each subject's prospective student profile report (PSPR) his predicted grade point average. These averages are a compilation of scores on the ACT in the areas of English, mathematics, social studies, and natural sciences and high school grades.

The predicted grade point average was used as the covariate rather than a composite percentile rank because it considers high school grades. Because the majority of the subjects were just out of high school, it appeared as if this would give the most accurate measure of their demonstrated academic ability.

Statistical Analysis

The study was primarily concerned with the significance of difference between the means of the several groups. Ferguson (1966) points out that analysis of variance is the proper statistical tool for rigorously testing for these differences. This type of analysis is based on the idea that whether or not the treatments applied have effect, some variation due to sampling fluctuation is still expected between means. If the variation cannot reasonably be attributed to sampling error, then the null hypothesis is rejected and the alternate hypothesis is accepted--that the treatments applied have an effect.

Another concern of the study was to ensure that the results observed could be attributed within limits of error to the treatment variable and to no other circumstance. Ferguson (1966) points out that a statistical, rather than an experimental method may be used to "control" or "adjust" for the effects of one or more uncontrolled variables, and permit, thereby, a valid evaluation of the outcome of the experiment. The analysis of covariance is such a method.

Sub-Design A. To test the significance of the differences between the four cueing levels and the two sex levels, a "four by two" analysis of covariance was run. The American College Test predicted grade point averages were used as the covariate or concomitant variable. To test hypotheses one and two, the main effects on the four cueing levels were examined. To test hypothesis three, the main effects on the two sex levels were examined. To test hypothesis five, interactive effects on the Cue and Sex factors were examined. Hypothesis four could not be tested for in this analysis.

Sub-Design B. A separate analysis of covariance was run for this design and the same covariate was used. The significance of the differences between the three cueing levels, the two sex levels, and the two relevance levels, was tested by using a "three by two by two" analysis.

To test further for hypothesis two, the main effects on the cueing dimension were examined. Hypothesis three was further tested by the main effects on the sex factor and hypothesis four by the main effects on the relevance factor. Hypothesis five was again tested for through examination of interactive effects between all three factors.

Level of Significance. The level of significance was set for both analyses at the .05 level.

FINDINGS

The purpose of this section is to present the results of the study. The results which follow have been separated into two sections. The first section presents those results with regard to Sub-Design A which dealt with hypotheses one, two, three, and five. The second section presents those results with regard to Sub-Design B which concerned hypotheses two, three, four, and five.

Sub-Design A

The first analysis was conducted to determine if the four groups under the eight treatment combinations differed significantly on the criterion test. Table 3 presents for the criterion test and the covariate the raw score means and standard deviations together with the number of subjects on which the descriptive statistics were calculated.

In spection of Table 3 reveals that the means of the several treatment combinations were quite dissimilar. The heterogeneity of the groups is further indicated by examination of the standard deviations. A covariate was used to decrease this heterogeneity, and Table 4 is presented to show the degree to which the several treatment combination means were modified. As Table 4 illustrates, reasonable homogeneity was established by this method.

Table 3. Raw score means criterion test and covariate.

Analysis A

Treatment abbreviation	Psychology in Life test			ACT predicted GPA		
	N	X	S	N	X	S
M-3	10	231.80	34.61	10	2.45	.36
M-2	15	203.47	37.98	15	2.18	.38
M-1	13	205.31	40.31	13	2.16	.32
M-N	9	216.78	38.30	9	2.32	.35
F-3	16	210.06	42.89	16	2.30	.39
F-2	17	215.82	47.63	17	2.37	.38
F-1	26	224.31	36.81	26	2.42	.30
F-N	28	220.36	33.03	28	2.40	.37
Total	134	223.41		134	2.34	

Table 4. Adjusted score means on criterion test

Cell abbreviation	N	X adj
Analysis A		
M-3	10	222.24
M-2	15	215.03
M-1	13	213.54
M-N	9	216.26
F-3	16	211.33
F-2	17	218.51
F-1	26	220.01
F-N	28	217.29
Total	134	216.76
Analysis B		
M-H-3	4	212.41
M-H-2	5	215.24
M-H-1	7	217.59
M-L-3	6	217.75
M-L-2	10	214.92
M-L-1	6	212.57
F-H-3	6	217.75
F-H-2	10	214.92
F-H-1	13	212.57
F-L-3	10	212.41
F-L-2	4	215.24
F-L-1	16	217.59
Total	97	215.08

An analysis of covariance conducted to determine if the differences between the treatment groups on the criterion test were statistically significant is presented in Table 5. Tests of the main effects were not significant as was true for the two-way interaction.

Table 5. Summary table for analysis of covariance on Sub-Design A

Source	df	Sum of squares	Mean squares	F
Cue	1	514.8	171.6	.1584
Sex	3	130.9	130.9	.1209
Cue-sex	3	1132.3	377.4	.3483
Regression	1	54007.4	54007.4	------
Error	125	135433.6	1083.5	------
Total	134	200196.0	------	------

F(3,125; .05)= 2.68 F(1,125; .05)= 3.92

The general interpretation to be drawn from the F-tests just reported is that the variables or factors introduced as cues had no significant impact on subsequent student achievement. Female susceptability was not demonstrated and little or no interaction between factors occurred.

Sub-Design B

The second analysis was conducted to determine if the three groups under the twelve treatment combinations differed significantly on the criterion test. Table 7 presents for the criterion test and the covariate the raw score means and standard deviations together with the number of subjects on which the descriptive statistics were calculated.

Table 6. Summary table for analysis of covariance on Sub-Design B

Source	df	Sum of squares	Mean squares	F
Cue	1	163.4	81.7	.0663
Sex	1	162.6	162.6	.1319
Relevance	2	312.5	312.5	.2535
Cue-Sex	2	834.8	417.4	.3386
Cue-Relevance	2	985.3	492.6	.3995
Sex-Relevance	1	165.7	165.7	.1344
Cue-Sex-Relevance	2	370.1	185.1	.1501
Regression	1	41,258.5	41,258.5	-----
Error	84	103,569.6	1,282.9	-----
Total	97			

$F(1,84\ .05) = 3.96$ $F(2,84\ .05) = 3.11$

Table 7. Raw score means criterion test and covariate

Analysis B

Treatment abbreviation	ACT predicted GPA			Psychology in Life Test		
	N	X	S	N	X	S
M-H-3	4	2.40	.32	4	218.75	14.84
M-H-2	5	2.26	.52	5	210.80	36.09
M-H-1	7	2.09	.34	7	205.14	39.19
M-L-3	6	2.48	.42	6	240.50	42.39
M-L-2	10	2.14	.32	10	199.80	40.26
M-L-1	6	2.25	.31	6	205.50	45.36
F-H-3	6	2.48	.42	6	214.17	55.46
F-H-2	13	2.37	.40	13	215.08	53.31
F-H-1	10	2.31	.23	10	214.10	32.26
F-L-3	10	2.19	.35	10	207.60	36.59
F-L-2	4	2.38	.38	4	218.25	26.86
F-L-1	16	2.48	.33	16	230.69	39.01
Total	134	2.32		134	224.52	

As was seen for Sub-Design A, the means on the criterion test for the several treatment combinations are very dissimilar even to a greater extent than in Sub-Design A. (Compare M-L-2 with M-L-3.) Inspection of Table 4 reveals the extensive homogenonizing effect of the covariate on the criterion means. As in Sub-Design A, the means for Sub-Design B were acceptably homogeneous after adjusting.

Another analysis of covariance conducted to determine if the differences between the treatment groups on the criterion test were statistically significant is presented in Table 6. Similar to the first analysis tests of the main effects were not significant. Neither of the three, two-way interactions nor the one, three-way interaction was significant.

The general interpretation to be drawn from these F-tests is that the variables or factors introduced as cues had insufficient effect o significantly facilitate subsequent student achievement. As in the first analysis, female susceptability was not demonstrated and little or no interaction took place. These F-tests also indicate that reported high or low relevance had insignificant influence on student achievement.

DISCUSSION OF FINDINGS

Introduction

The essential purpose of this research project was to determine the feasibility of establishing a classroom experimental situation in which subject-object and subject-observer cues would be introduced to determine if such cues had any subsequent effect on student achievement as measured on the criterion tests to indicate the Hawthorne effect. Concurrent with this feasibility study was an attempt to determine if the presence of such cues operated in a manner such that female achievement would be heightened as compared to male achievement. Combined with these two purposes was a third purpose which focused upon the question of whether or not perceived relevance in terms of the experimental situation significantly influenced achievement. The fourth purpose was to discover new information concerning interaction between experimental factors.

The general results from the two separate analyses indicated that various combinations of subject-object and subject-observer cues did not operate in a manner such that achievement was significantly affected. In other words, there was no Hawthorne effect. Further, no significant differences were observed in terms of sex or relevance. Interaction, if it existed, was not operating to any recognizable degree.

Limitations

A review of the general experimental design employed plus an examination of specific procedures suggests several possibilities as to why the results reported above may have occurred.

To parallel very closely Cook's (1967) explanation, possibly no such variable known as the Hawthorne effect exists at least as a variable of powerful enough influence to influence significantly any psychological or educational study.

Regarding the design itself there were many indications from answers on the students' questionnaires and comments made to the assistant instructor that the cues were of sufficient sophistication to produce the aura of special experimental procedures. A majority of responses from the student questionnaires indicated that they felt the situation was real although not relevant to them individually. Many indicated that there were hidden microphones and observers. This seems to reveal their acceptance of the cues as authentic. It was reported to the investigator that several male students in the Cue-1 group actually entered a large air vent in the room to search for hidden cameras, microphones, and observers. This data seems to argue against the idea that cueing was insufficient or inadequate, however, this possibility should be considered.

In conjunction with the above consideration is the question of ego involvement. Whereas in the original Hawthorne studies the subjects' lives were being modified by the experimental treatments thus assuring a high degree of ego involvement, in this study no such far reaching effects of treatments were employed. The subjects were informed that the study would not influence their grades, so in retrospect it appears that if there existed any potentially strong link between the ego involvement and experimental treatments, it was broken.

Cook (1967, p. 101) discusses cognitive and affectional awareness with the inference that simple cognitive awareness may not be sufficient to elicit the Hawthorne effect. Judging by the behavior described above and impressions written on the questionnaire, it appeared that a good deal of affective awareness accompanied the cognitive awareness in this study.

The length of the study may be an explanation for lack of any measureable Hawthorne effect. The students were only exposed to cueing for a period of nine weeks, whereas, in the original Hawthorne studies, treatments went on for months at a time.

Another explanation might be that students by the time they reach university level are desensitized to the effects of any new innovations. This may be true because of the ever increasing frequency of progressive teaching methods and audio-visual materials that accompany students as they progress through the various academic levels.

Another explanation might center around the criterion instrumentation, the general psychology test, which may not have been sufficiently valid measure of the actual instruction so that it would reflect achievement gains initiated by the cues that were employed. As Cook (1967) points out, the criterion instrumentation problem is not unique to a study of this type since it is present in almost all attempts to evaluate educational innovations. The same explanation might be applied to the Relevance scale also.

A further possible explanation might be that the Hawthorne effect acted on all groups equally. This is a common explanation made by educational and psychological researchers when expected differences between control and experimental groups fail to appear. Nevertheless, several students from the control group indicated on their questionnaire when asked if they had known an experimental situation existed answered, "Yes," contrary to what was expected. Their answers revealed that the teaching assistant had administered a Kuder Interest Inventory and a questionnaire of her own making to all laboratory groups. This had been construed by several students to mean that they were under experimentation. It is difficult to say just how potent these unfortunate irregularities might have been.

To parallel again another of Cook's (1967) explanations, possibly, awareness of the Hawthorne effect and student achievement are two variables which do not have any fundamental relationship. Awareness of experimentation may be one human attribute and achievement gain another, and to posit a relationship between them in the sense that awareness of Hawthorne effect is an independent variable and achievement a dependent variable may not be reasonable.

Another explanation is that from all appearances the most potent motivation for academic achievement, particularly at an undergraduate level, is teacher evaluation. Because the three experimental groups were informed the first laboratory period that the study would have no effect on their grades, possibly this obliterated the link between environment manipulation and motivation.

Another important possibility that should be considered is personality factors. Such factors as susceptibility, rigidity, dogmatism, introversion, etc., might well have played a crucial role in the differential creation of the Hawthorne effect. Although the personality factor of susceptibility was considered to some degree in terms of sex, still it might have been pursued further. If an adequate measure of susceptibility could have been created and then used as another factor in the factorial design, discovery of underlying dynamics might well have been facilitated.

The final consideration of limitations of this study is the kind of Hawthorne effect whose creation was being attempted. Surely a difference must exist between the type of limited effect that was sought in this study and the effect that could result from a new program which is advertised in newspapers and over radio and television, that included parents and school staffs as well as the students under awareness of special experimental treatment. Modification of the total environment might well produce a type of Hawthorne effect completely foreign to the type attempted in the few hours of cueing presented in this study.

Implications

The purpose of this section is to set forth what appears to be the principal implications of the research effort with regard to the nature and function of the Hawthorne effect concept in educational and psychological research. Several principal implications appear to be justifiable in terms of the results and conclusions presented above. Because the results were similar to Cook's (1967) results, the implications run parallel also.

The first implication is that perhaps much of what has been written about the nature, operation, and control of the Hawthorne effect in psychological and educational research appears to have been generated largely on the basis of intuition and logic rather than upon any empirical basis.

If this study which was set up to create purposively the effect, failed, then how can it justifiably follow that such an effect will inadvertently act at an extremely obscure level yet significantly and distortingly affect research? To quote Cook (1967, p. 130), "The Hawthorne effect concept is being put in the position of being guilty (i.e., operational) with efforts then being directed to establish its innocence rather than being considered as innocent (i.e., nonoperational) until its guilt has been established." In short, the existence of the Hawthorne effect needs to be adequately and empirically demonstrated.

The second implication is that if researchers continue to accept the Hawthorne effect as an actual dynamic and define it in somewhat the same way as Cook (1967) does, then it must be decided whether simple awareness of experimental or special circumstances can be equated with the Hawthorne effect. Research is needed to tell us whether it is conscious or unconscious, whether adults and children react the same to it, and whether in fact it even needs to be controlled.

The third implication is that based on this and Cook's study, researchers do not need to control for this effect. By the same token researchers can not justifiably blame the Hawthorne effect for lack of significant differences. They may just have to admit that their new methods or research designs are not adequate enough to produce significant differences.

SUMMARY

The general purpose for this study was to discover if the Hawthorne effect could be experimentally created in a freshman general psychology class as measured by the criterion tests. If the effect was created, then it was the purpose of this study to discover the influence of subject-object, subject-observer, and subject-object-observer cues; the influence of sex; the influence of perceived relevance; and the influence of interaction between the experimental variables on the experimental creation of the Hawthorne effect.

The subjects for this study were 134 Utah State University students who during fall quarter, 1967, were enrolled in a general psychology course.

A post test only, control group design was employed. Each week for nine weeks, three of these four groups of subjects were exposed to different cues all of which were intended to create the aura of experimental conditions and special treatment.

The instrumentation was accomplished through administering six multiple-choice tests on general psychological principles and one questionnaire. The six scores from the tests were summed to give one criterion score for each subject. The questionnaire was examined for biographical information and manifest relevance.

The statistical analysis was comprised of two analyses of covariance. The covariates were predicted grade point averages as derived from each subject's ACT scores.

Because the hypotheses represented the culmination of the problems and purposes, each will be considered separately with the applicable conclusions.

Hypothesis One

The experimental groups will show greater influence from the Hawthorne effect than the control group.

As indicated by the F-tests in the analysis for Sub-Design A, the tests of the main effects were not significant as was true for the two-way interaction. It was concluded from this that hypothesis one was not realized. The experimental groups did not show greater influence from the Hawthorne effect than the control group. It may be concluded that neither experimental nor control group showed Hawthorne effect, but this did not necessarily follow. The possibility remains that both could have come under the effect.

Hypothesis Two

Within the three experimental groups, there will be an increasing Hawthorne effect with the least effect in the subject-object cue group and the greatest effect in the subject-object-observer group as compared to the control group.

The F-tests in both Sub-Design A and B analyses indicated that there were no significant differences between these three experimental groups. Hypothesis two

was not realized. The conclusion is that either the cues were too weak to elicit the effect or that more subtle cues were adequate and further elaboration was superfluous, therefore, no differentiation between groups by cueing.

Hypothesis Three

The females in all experimental groups will show significantly greater Hawthorne effect than males within the same groups.

Both analyses of covariance failed to support this hypothesis. The conclusion was that females under these conditions were not more susceptable, and therefore, not more highly motivated.

Hypothesis Four

The group rating high on the Relevance scale will show significantly greater Hawthorne effect than the groups rating low in relevance.

Based on the several F-tests from both analyses it was concluded that those subjects who indicated that they thought the experimental situation was relevant to scientific exploration did not react more strongly to the Hawthorne effect than those who did not think it relevant. Hypothesis four was not realized. It was also concluded that relevance as defined in this study is not related to differential student achievement.

Hypothesis Five

There will be sufficient interactive effects between factors to the extent that some will reach significance.

Interaction between factors was not demonstrated by either of the analyses of covariance. It was hoped that this might be a fruitful avenue to gaining information about the underlying dynamics of the effect under study, but it was concluded that the dynamics were not isolated to a sufficient enough degree to be useful.

In way of summary, the general objective of the study was only partially realized. Since none of the hypotheses were demonstrated, the conclusion follows that to create experimentally the Hawthorne effect, different procedures must be followed than were used in this study. To the degree that the results of this study can be generalized, it would seem that no such effect exists.

BIBLIOGRAPHY

Borg, W.R. 1963. Educational Research: An Introduction. New York: David McKay and Company.

Campbell, D.T. and Stanley, J.C. 1963. "Experimental and quasi-experimental designs for research on teaching." In N.L. Gage (Ed.) Handbook of Research on Teaching. Chicago: Rand McNally.

Clark, H.F. 1963. Cost and Quality in Public Education. Syracuse, New York: Syracuse University Press.

Cook, D.L. 1962. "The Hawthorne effect in educational research." Phi Delta Kappan. 44; 116-122.

Cook, D.L. 1967. The Impact of the Hawthorne Effect in Experimental Designs in Educational Research. Final Report, Project No. 1757, U.S. Office of Education. June.

Cronbach, L. 1964. "The evaluation for course improvement." New Curricula. New York: Harper and Row.

Dixon, T.R. 1966. "The effects of experimenter-subject relationship upon verbal habits and verbal conditioning." The Journal of General Psychology. 75; 151-156.

Ferguson, G.A. 1966. Statistical Analysis in Psychology and Education. New York: McGraw Hill.

Fillenbaum, S. 1966. "Prior deception and subsequent experimental performance: The 'faithful' subject." Journal of Personality and Social Psychology. 4 (5); 532-537.

Frank, J.D. 1944. "Experimental studies of personal pressure and resistance: I. Experimental production of resistance." Journal of General Psychology. 30; 23-41.

Gall, M., and Mendelsohn, G.A. 1967. "Effects of facilitating techniques and subject-experimenter interaction on creative problem solving." Journal of Personality and Social Psychology. 5 (2): 211-216.

Hurst, Rex L. 1967. General Least Squares Computer Program. Utah State University Computer Center. (Mimeographed)

Landesberger, H. 1958. Hawthorne Revisited. Ithaca, New York: Cornell University Press.

Mayo, Elton. 1946. "The Hawthorne Experiment, Western Electric Company." The Human Problems of an Industrial Civilization. Boston, Mass: Division of Research, Graduate School of Business Administration, Harvard University.

McCall, W.A. 1923. How to Experiment in Education. New York: Macmillan.

Orne, M.T. 1962. "On the social psychology of the psychological experiment; with particular reference to demand characteristics and their implications." American Psychologist. 17: 776-783.

Pennock, G.A. 1929. "Industrial research at Hawthorne: An experimental investigation of rest periods, working conditions, and other influences. The Personnel Journal. 8: 296-313.

Riechen, H.W. 1962. "A program for research in experiments on social psychology." In N.T. Washburne (Ed.) Decisions, Values, and Groups. Vol. 2. New York: Macmillan.

Roethlisberger, F.J., and Dickson, W.J. 1947. Management and the Worker. Cambridge, Massachusetts: Harvard University Press.

Ruch, F.L., Warren, N., and Gorfein, D.S. 1967. Psychology and Life. 7th Edition, Chicago, Illinois: Scott Foresman & Company.

Shaver, J.P. 1967. From lecture notes of Educational Research course (Ed 267). Recorded November, 1967, by this writer at Utah State University.

Snow, C.E. 1927. "Research in illumination." Tech. Engineering News. 8: 257-282.

Travers, R.M. 1958. An Introduction to Educational Research. New York: Macmillan.

Tyler, L.W. 1965. The Psychology of Individual Differences. 3rd Edition. New York: Appleton-Century-Crofts.

APPENDIX

QUESTIONNAIRE

PSYCHOLOGICAL INQUIRY

This questionnaire is an important part of a psychological study being conducted in conjunction with Psych. 53. All information asked for is pertinent to the study and will be held as confidential.

Name (print) _____ Lab Section _____

---

1. Did you know before seeing this questionnaire that you were part of a psychological study?

(Circle one) YES NO

2. Approximately when did you become aware that you were under some kind of research?

(Circle one) (weeks in quarter)

1st 2nd 3rd 4th 5th 6th 7th

8th 9th 10th

3. What indications had you that you were under study?

---

4. What do you feel is the experimenter's reason for conducting a psychological study on this class?_____

---

5. Do you feel you were observed by any other means than were obvious during class?

(Circle one) YES NO

(If YES, which one or ones?) Hidden microphone, camera, observer, etc.

6. Do you feel that being under experimental conditions has influenced your performance in this class?

(Circle one) YES NO

(If YES, indicate which way-- Postively or Negatively)

APPENDIX (continued)

7. What do you think the experimenter's hypothesis is concerning the outcome of this study?_____

_____

_____

8. Do you feel the experimenter's hypothesis was realized?

(Circle one) YES NO

Explain:_____

_____

9. On the provided scale, mark at the appropriate region your feeling about the degree of importance that this study is to the advancement of science in the area of human psychology.

A definite waste of time and energy — Neither here nor there — Highly significant

(-)_____(+)

Could use time better — A step in the right direction

10. Did you take this class because it was required?

(Circle one) YES NO

11. What is your major?_____

12. What is your class at the university?

(Circle one) Fr. Soph. Jr. Sr.

13. What is your age?_____

14. Where is your home town? _____ City _____ State

APPENDIX (continued)

15. Why did you sign up for this particular laboratory section?

(Circle one)

a. It was the only one available.

b. It was the only one which did not conflict with other classes.

c. I like this time of day best.

d. I had friends in the same section.

e. Other reasons (Explain)_____

_____

VITA

Richard Carl Harris, Jr.

Candidate for the Degree of

Master of Science

Thesis: Differential Hawthorne Effect by Cueing, Sex, and Relevance

Major Field: Psychology

Biographical Information:

Personal Data: Born at Logan, Utah, July 10, 1940, son of Richard C. and Margaret Crowther Harris; married Melanie Fox, August 27, 1964; one child--Heidi.

Education: Attended elementary school in Elwood, Utah; graduated from Bear River High School in 1958; received Bachelor of Science degree from Utah State University with a major in psychology and a minor in zoology, 1966-67; completed requirements for the Master of Science degree specializing in psychology at Utah State University in 1968.