

AN INVESTIGATION OF EXAMINER
INFLUENCE ON WECHSLER INTELLIGENCE
SCALE FOR CHILDREN SCORES

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
William Harvey Gillingham
1970



This is to certify that the
thesis entitled
An Investigation of Examiner
Influence on Wechsler Intelligence
Scale for Children Scores

presented by

William Harvey Gillingham

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Counseling, Personnel
Services and Educational Psychology

Robert C. Craig
Major professor

Date 1/29/70

~~JUN 9 1974~~ ~~2-18-92~~
~~16167~~ ~~227~~ ~~10-10-74~~ ~~102~~

~~AUG 3 1976~~ ~~235~~
~~SEP 28 1975~~ ~~231~~
~~OCT 1 1975~~ ~~181~~
~~AUG 27 1984~~ ~~227~~

ABSTRACT

AN INVESTIGATION OF EXAMINER INFLUENCE ON WECHSLER INTELLIGENCE SCALE FOR CHILDREN SCORES

BY

William Harvey Gillingham

A. The Problem

Discussion of the possibility of observer influence has recently become controversial largely as a result of the work of Robert Rosenthal (1966; 1967a). Observer influence has been examined by studying different types of observers such as interviewers, research experimenters, teachers, and projective and intelligence test examiners. This thesis was a study of examiner influence in that it looked at the person who administered the Wechsler Intelligence Scale for Children (WISC) as a possible source of undesired variation of intelligence test scores.

B. Design

The sample consisted of four male and four female WISC examiners who had just completed WISC training. Each examiner tested eight junior high school students (four boys and four girls) who were randomly selected from a population of students who were "average" in intelligence (that is, they scored from ninety to one hundred-ten on the California Test of Mental Maturity-Short Form, 1963 Revision).

Eight examinees were randomly assigned to each examiner and then they were randomly designated "above average" or "below average" so that each examiner tested two "above average" boys, two "below average" boys, two "above average" girls, and two "below average" girls. The order in which an examiner tested his examinees was also randomized. Parental approval for testing was secured and confidentiality was assured. Neither parents nor examiners were told that they were participating in a study or an experiment. All WISC's were administered in private offices during a one week period.

A rating sheet was used by the grand experimenter to transmit an expectancy condition to examiners. On each examinee's rating sheet a discrepancy between California Test of Mental Maturity-Short Form score and school achievement was fictitiously indicated, and a "predicted WISC score" was advanced.

Analysis of the data was by a mixed model, four-way analysis of variance having three fixed variables (expectancy, sex of examinee, and sex of examiner), and two random variables (examiners nested in sex of examiner, and replications nested in all other variables). The five per cent level of confidence was arbitrarily chosen for significance tests.

C. Analysis of Results

Both the sex of examiner and sex of examinee effects were significant ($p < .05$). Female examiners obtained higher mean WISC scores and male examinees achieved higher mean WISC scores.

The expectancy effect was not statistically significant. It was feared that WISC examiners did not retain the expectancy condition

given to them by the grand experimenter and, therefore, examiners' expectancy was neither transmitted to nor received by examinees. It was also felt that the WISC was a structured and factual experimental task and that the WISC examiners were relatively well-trained and experienced. Neither the examiner effect nor any of the interaction effects was significant.

The results of this study pointed to the conclusion that examiner influence was not a great problem in intelligence testing and that observer influence was difficult to demonstrate when experienced observers administered structured tasks.

AN INVESTIGATION OF EXAMINER INFLUENCE ON WECHSLER
INTELLIGENCE SCALE FOR CHILDREN SCORES

By

William Harvey Gillingham

A THESIS

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

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services
and Educational Psychology

1970

ACKNOWLEDGEMENTS

The writer is especially indebted to the late Buford Stefflre under whose direction this thesis originated. Dr. Stefflre's impact on the writer's life is continually felt. Special gratitude is also felt for Dr. Robert Craig who kindly and patiently assumed an advisory role, and to other committee members, Dr. Donald Grummon and Dr. Fred Vescolani.

The writer would also like to acknowledge his wife, Judy, for her support, understanding, and unselfishness throughout his doctoral program.

TABLE OF CONTENTS

CHAPTER	PAGE
I. THE PROBLEM	1
Need for the Study	1
Purpose of the Study	3
Hypotheses	4
Definition of Terms	6
Overview of the Study	7
II. REVIEW OF RELATED RESEARCH	8
Examiner Influence	8
Experimenter Influence	11
Discussion of Previous Research	23
III. DESIGN	26
The Sample	26
Instrumentation	27
Procedure	30
Design and Analysis	31
Summary	35
IV. ANALYSIS OF RESULTS	37
Hypotheses and Results	37
Discussion of Results	42
Summary	49

CHAPTER	PAGE
V. SUMMARY, CONCLUSIONS, AND IMPLICATIONS FOR FURTHER RESEARCH	52
Summary	52
Conclusions	59
Implications for Further Research	62
BIBLIOGRAPHY	64
APPENDIX A. TABLE 6. CTMM MEANS AND STANDARD DEVIATIONS FOR ALL SUBGROUPS	71
APPENDIX B. Rating Sheet	73
APPENDIX C. TABLE 7. WISC TOTAL SCORES	74

LIST OF TABLES

TABLE	PAGE
1. SOURCES OF VARIATION AND DEGREES OF FREEDOM	34
2. SUMMARY OF MEAN TOTAL WISC SCORES, STANDARD DEVIATIONS, AND F-RATIOS FOR MAIN EFFECTS	38
3. SUMMARY OF MEAN TOTAL WISC SCORES, STANDARD DEVIATIONS, AND F-RATIOS FOR TWO-WAY INTERACTION EFFECTS	40
4. SUMMARY OF MEAN TOTAL WISC SCORES, STANDARD DEVIATIONS, AND F-RATIO FOR THREE-WAY INTERACTION EFFECT	42
5. SUMMARY OF MEANS, STANDARD DEVIATIONS, AND F-RATIOS FOR SIGNIFICANT EFFECTS	58
6. APPENDIX A. GTMM MEANS AND STANDARD DEVIATIONS FOR ALL SUBGROUPS	71
7. APPENDIX C. WISC TOTAL SCORES	74

LIST OF FIGURES

FIGURE	PAGE
1. Research Design	32

CHAPTER I

THE PROBLEM

In this chapter the need for the study is developed and then the purpose of the study is specifically stated. The research hypotheses are stated and certain terms are defined. The chapter is concluded with an overview of the study.

A. Need for the Study

The possibility that observers may influence what they observe or may be influenced in how they record what they observe has been discussed for years (Kintz, Delprato, Mettee, Persons, & Schappe, 1965). Discussion of the possibility of observer influence has recently become controversial largely as a result of the work of Robert Rosenthal (1966, 1967a). Observer influence has been examined by studying different types of observers. Thus, Rosenthal (1966, 1967a) studied the research experimenter to determine the existence of "experimenter influence." Rosenthal and Jacobson (1966, 1968) examined the possibility of "teacher influence" by attempting to see if teachers influenced pupils' scores on intelligence tests. Sociologists (Suchman, 1962; and Wilkie, 1963) looked at the sociological interviewer as an observer to determine the possibility of an "interviewer influence" effect. The possibility of "examiner influence" has been checked by studying the examiners of individual intelligence

tests (Cleutat, 1965; Cohen; and Larrabee and Kleinsasser, 1967), and by studying the examiners of projective tests (Simmons and Christy, 1962; Turner and Coleman, 1962; and Rosenthal, 1963b).

Rosenthal (1966; 1967a) is convinced that observer influence (particularly experimenter influence and teacher influence) exists. He claims to have repeatedly demonstrated experimenter influence in experimental studies. His data and their interpretation have been sharply criticized (Ingraham and Harrington, 1966, 1967; Barber and Silver, 1968a, 1968b; Thorndike, 1968; Barber, Forgione, Chaves, Calverley, McPeake, & Bowen, 1969; and Claiborn, 1969). Some critics concede the possibility of experimenter influence but question the pervasiveness of observer influence in more structured situations such as individual intelligence testing. Other criticisms center around shortcomings in Rosenthal's research designs and his statistical treatment of data.

The existence of observer influence needs to be examined in other than experimenter influence studies. An area of concern for the writer has been the use of individual intelligence tests to make important and sometimes irreversible decisions about people. If this practice continues it seems imperative that the possibility of examiner influence be investigated. Furthermore, future studies of observer influence of all kinds need to be statistically well-designed so that they do not meet with the same criticisms leveled at Rosenthal. In short, there seems to be a need for a well-designed study to investigate the existence of examiner influence on individual intelligence test scores.

B. Purpose of the Study

The purpose of this study was to provide a well-designed field testing of experimental research results. Specifically, this study looked at the person who administered the Wechsler Intelligence Scale for Children as a possible source of undesired variation of intelligence test scores. The research design also provided for an examination of variables that had been rarely isolated in previous research. In this design, test examiners were given both expectancy conditions in random order; that is, examiners were led to believe that half of their examinees were "above average" in intelligence and half were "below average" in intelligence. The sex of examiner and sex of examinee variables were crossed so that male examiners tested male and female examinees and female examiners tested male and female examinees. The examinees were junior high school students and therefore, were much younger than the examiners who were graduate students at Michigan State University.

C. Hypotheses

Eight hypotheses were generated from the design of this study.

Hypothesis 1

Female examiners will obtain a higher mean intelligence test score than male examiners. Some research (Cieutat, 1965) suggests that female examiners of individual intelligence tests obtain higher scores from examinees.

Hypothesis 2

There will be no difference between mean intelligence test scores achieved by boy and girl examinees. No research was found that would suggest that male examinees score higher or lower than female examinees on the Wechsler Intelligence Scale for Children.

Hypothesis 3

Examinees who are expected by examiners to be above average will achieve a higher mean intelligence test score than examinees who are expected to be below average. This hypothesis is a test of the expectancy effect which has been experimentally demonstrated (Larrabee and Kleinsasser, 1967; Rosenthal, Fode, Friedman, and Vikan-Kline, 1960; Rosenthal, Persinger, Vikan-Kline, and Mulry, 1963; Rosenthal, Mulry, Persinger, Vikan-Kline, and Grothe, 1964a, 1964b; and Friedman, Kurland, and Rosenthal, 1965). This hypothesis will be directionally tested.

Hypothesis 4

There will be no reliable difference between mean intelligence test scores obtained by the examiners. Since examinees were randomly assigned, there is no reason to expect that examiners will obtain different mean intelligence test scores. If this test is statistically significant, it probably indicates that some examiners regularly assign scores that are higher than scores assigned by other examiners.

Hypothesis 5

Opposite sex combinations of examiners and examinees will

obtain higher mean intelligence test scores than same sex combinations. There is sufficient research (Stevenson, 1961; Stevenson and Allen, 1964; Cieutat, 1965; Hill and Stevenson, 1965; and Rosenthal, 1967a) to support a directional hypothesis concerning the sex of examiner interacting with the sex of examinee variable.

Hypothesis 6

There will be no interaction effect between the sex of the examiner and the extent to which the obtained scores are in the direction suggested by the expectancy set. It cannot be determined whether male or female experimenters are more susceptible to an expectancy effect. Rosenthal, Mulry, Persinger, Vikan-Kline, and Grothe (1964a) found that male experimenters are "better biasers," but a directional hypothesis is not advanced.

Hypothesis 7

There will be no interaction effect between the sex of the examinee and the extent to which the obtained scores are in the direction suggested by the expectancy set. Again, a directional hypothesis is not advanced because of insufficient research to support the conceptualization that one sex of examinees is more influenced by the direction of the expectancy than the other sex of examinees. In two studies (Rosenthal, Mulry, Persinger, Vikan-Kline, and Grothe, 1964a; and Stevenson and Allen, 1964), it was found that female subjects were easier to influence than male subjects. As in Hypothesis 6, a directional hypothesis is not formulated because it is felt that the subjects' age may have accounted for the results found in the studies

cited. That is, a young male examiner may be better able to influence a college coed than an eleven year old junior high school girl.

Hypothesis 8

There will be no interaction effect among the sex of examiner, sex of examinee, and direction of expectancy variables. This hypothesis is advanced to determine the existence of a three-way interaction effect.

D. Definition of Terms

The following terms need to be defined to assure common meaning.

Expectancy

Experimenters may have a preconceived idea concerning the results of their research. A test examiner may have a preconceived idea concerning the test score of the subject he is testing. That is, he has an "expectation" of what the outcome will be. For the purpose of the study, an "expectancy" was believed to exist when the grand experimenter (the writer) gave test examiners an estimate of the ability of each subject tested.

Experimenter of Examiner Influence

"Experimenter or examiner influence" involves a conscious or unconscious effect on the outcome of research or test scores.

E. Overview of the Study

In the next chapter, a review of recent (1960 to present) research on experimenter and examiner influence is presented. The

reader will be introduced to the research that has been done on the experimenter and examiner variables. He will also be shown that the present study represents a necessary extension of the previous research.

The research design (sample, instruments, procedure, and analysis) is discussed in Chapter III and results are presented in Chapter IV. Chapter V includes a summary, conclusions, and implications for further research.

CHAPTER II

REVIEW OF RELATED RESEARCH

Chapter II contains a review of recent (1960 to present) literature concerning examiner influence on individual intelligence test scores and on projective test scores, research on experimenter influence, and a discussion of previous research. Studies of examiner influence on individual intelligence and projective test scores were included because of their relevance to this study's design in which an individual intelligence test was used. These studies were conducted to determine whether physical or personality characteristics of test examiners influenced the test scores they obtained.

The rest of the studies in Chapter II were concerned with Robert Rosenthal's (1967a) contention that experimenters influence the results of their research.

A. Examiner Influence

Examiner Influence on Individual Intelligence Test Scores

Very little research was found concerning examiner influence on intelligence test scores. Cieutat (1965) also noted this lack of research. In his study, seven male and six female examiners tested 243 boys and girls with the Stanford-Binet. Cieutat found that female examiners obtained significantly higher ($p \leq .001$) intelligence test scores from subjects. Analysis of variance revealed a

significant interaction effect ($p \leq .05$) between sex of examiners and sex of subjects. Highest performances were obtained when examiners tested opposite sex subjects. Cohen (1965) also found a significant ($p \leq .005$) examiner effect using the Wechsler-Bellevue. Cohen believed that examiner influence reduced the subtest validity of the Wechsler-Bellevue.

Exner (1966) studied the effect of "rapport building" on Stanford-Binet intelligence test scores. Twenty-five subjects took the test after substantial attempts were made by examiners to "build rapport." The other twenty-five examinees took the test without preliminary "rapport building." The experimental group scored significantly higher ($p \leq .001$) than the control group.

Wartenberg-Ekren (1962) found no significant difference on Block Design scores of the Wechsler Adult Intelligence Scale when eight examiners gave that subtest to two examinees who were allegedly earning higher grades than the other two examinees. However, Larrabee and Kleinsasser (1967) found a significant difference when five examiners administered the Wechsler Intelligence Scale for Children to twelve sixth graders. Each examinee was tested by two examiners: one examiner administered the even items; the other examiner administered the odd items. One examiner was told that the examinee was "above average"; the other examiner was told that the examinee was "below average." On the Verbal part of the test the difference was over ten IQ points ($p \leq .05$).

Friedman (1967) reviewed literature concerning examiner differences in testing and found that the examiner's and subject's race made a difference in testing. Same race combinations obtained best

scores. He noted, however, the lack of research on the examiner variable in testing. He pointed out that most subjects were still tested by only one examiner. He also suspected an examiner expectancy effect in testing because most examiners review the examinee's records before testing.

Examiner Influence on Projective Test Scores

Turner and Coleman (1962) and Simmons and Christy (1962) studied examiner influence on Thematic Apperception Test (TAT) responses. Turner and Coleman designed their study to maximize the probability of significant differences but failed to obtain many significant results. They did find that examiners who exhibited warmth elicited significantly more hostile responses from subjects. They did not find examiner competency, experience, or preference for the TAT significantly related to subjects' responses.

Rosenthal (1963b) reported two studies conducted in the 1950's in which subjects' scores on Rorschach cards correlated significantly with examiners' scores. Furthermore, analysis of subjects' pretest and posttest scores revealed that their score became more like the examiners' scores as a result of the examiner-subject interaction.

Examiner influence on individual intelligence test scores or on projective test scores was investigated in a limited number of research studies. In these studies, the examiner's personality (warmth, rapport) and physical characteristics (sex, race) had an effect on the dependent variable. Specifically, opposite sex and same race combinations obtained the best scores; "warm" examiners elicited more hostile responses; and examiners who built rapport

obtained significantly higher Stanford-Binet intelligence test scores. Examiner influence on intelligence and projective test scores prevents accurate assessment of examinees and therefore, needs to be more fully understood and controlled.

B. Experimenter Influence

There were many studies of experimenter characteristics and their influence on experimental research. These included studies of the sex of experimenter, prestige of experimenter, experience of experimenter, personality of experimenter, and race of the experimenter. Several studies were reviewed in which the cues which mediate experimenter influence were examined.

Sex of Experimenter

None of the twelve studies reviewed in this section demonstrated a significant sex of experimenter effect. Two of seven studies showed a significant sex of subject effect--in one study male subjects performed better and in the other study female subjects performed better. When the sex of experimenter variable was examined in interaction with another variable, significant results were usually obtained. The interaction of the sex of experimenter and sex of subject variables was sometimes significant--especially when experimenters worked with opposite sex subjects.

Stevenson and Allen (1964) found significant sex of subject and interaction effects when adult subjects were verbally reinforced in performing a marble-sorting task. Female subjects performed significantly better than males and opposite sex interactions resulted

in the best performances. The authors (Stevenson and Allen, 1967) repeated this study later, however, and found no significant interaction between the experimenter and sex of subject variables.

Miller and Solkoff (1965) did not find a significant sex of experimenter effect in a verbal conditioning experiment. Stevenson (1961) studied children with the marble-sorting task and found sex of experimenter related to the age and sex of subjects. Opposite sex combinations of experimenters and subjects resulted in the best performances. In another study (Hill and Stevenson, 1965) using marble-sorting as the dependent variable, no significant sex effects were found until the interaction of sex and verbal reinforcement was examined. Best performances were obtained by male experimenters in the reinforcement condition.

A verbal conditioning study (Sarason and Minard, 1963) revealed that male subjects were conditioned easier than female subjects, but sex of experimenter was not a significant variable. The interaction effect of the sex of experimenter and hostility of experimenter variables was significant. Low hostility male experimenters and high hostility female experimenters conditioned their subjects best. In another verbal conditioning experiment (Ogawa and Oakes, 1965) it was found that female experimenters conditioned low anxiety male subjects significantly better than high anxiety male subjects. The authors reasoned that high anxious males became more anxious which reduced the quality of their performances.

A summary of these verbal conditioning studies indicated that the sex of an experimenter alone did not influence the results of the studies. Only when another independent variable interacted with the

sex of experimenter variable were significant results obtained.

Prestige of Experimenter

Four studies were reviewed which dealt specifically with the prestige of experimenters as a variable. All of these studies involved verbal conditioning. Two of the studies revealed significant prestige effects; the other two did not. Of the former two studies, one showed that low prestige experimenters obtained better performances; the other study reported that high prestige experimenters obtained better results. These studies were discussed below.

Prince (1962) found that more prestigious experimenters conditioned the verbal behavior of children significantly better than less prestigious experimenters. In another verbal conditioning experiment (Katkin, Risk, and Spielberger, 1966) an undergraduate experimenter (low prestige) obtained a significantly greater performance increment than a professor (high prestige). With only two experimenters it was possible that other variables may have influenced the subjects' responses in this experiment.

Sarason and Minard (1963) manipulated the experimenters' prestige level by the way they dressed and contacted subjects. They found no significant prestige of experimenter effect using sixteen experimenters in a verbal conditioning experiment. Blaufarb (1960) found no significant prestige of experimenter effect using ten experimenters.

In a social science survey research study, Jones (1963) discussed a "courtesy bias" found in Southeast Asia. She referred to the experimenter effect of prestige on people who traditionally treated

visitors courteously, and stressed the importance of constructing questionnaire items which were not susceptible to influence.

Experience of Experimenter

The effect of an experimenter's experience on research results was studied by several authors. Rosenthal (1964) once suggested that use of unsophisticated and less ego-involved experimenters might reduce an experimenter's influence. But he found that naive experimenters influenced results also and retracted his former suggestion. In an animal conditioning experiment, Brogden (1962) found that naive experimenters did not condition rabbits as well as experienced experimenters. However, the difference diminished with practice. Cordare and Ison (1963) reported an interesting study of planaria. Experimenters who expected high planarian activity reported significantly ($p \leq .001$) more activity than those experimenters expecting little planarian activity. They attributed this extremely significant difference to the naïveté of the experimenters. Another study (Ingraham and Harrington, 1966) appeared to support this notion. Some of the experimenters worked with rats which they were led to believe were "bright"; others worked with rats which were supposedly "dull." They found no significant difference and even the insignificant difference disappeared by the fifth day of the experiment. The authors conceptualized that naive experimenters must rely on cues given by the chief experimenter because of the ambiguity of the situation. As experimenters gained experience in the experimental situation, they increasingly responded to factual cues emitted by the rats. The authors concluded that experimenter influence was not a problem if experimenters were experienced in

working with subjects. In a sociological study, Kish (1962) also found that experienced interviewers influenced their data significantly less than naive interviewers. He suggested the use of structured questionnaire items to reduce interviewer influence.

In three of four studies, naive experimenters influenced their results more than experienced experimenters and this influence was reduced as experimenters practiced or became more experienced.

Personality Characteristics of Experimenter

Many studies were reviewed which examined the effect of experimenters' personality in experiments. Variables included hostility, anxiety, need for social approval, and warmth of the experimenter.

Hostility. Two studies (Sarason, 1962; and Sarason and Minard, 1963) demonstrated a significant experimenter hostility effect on the results of verbal conditioning experiments. Hostile experimenters of both sexes elicited more hostile verbs in one study (Sarason, 1962). In the other study, low hostility male experimenters and high hostility female experimenters were able to condition subjects best.

Anxiety. In discussing the effect of experimenters' anxiety on results, researchers (Rosenthal, Persinger, and Fode, 1962; Winkel and Sarason, 1964) felt that a curvilinear relationship existed. That is, experimenters exhibiting medium anxiety influenced subjects more than experimenters showing high or low anxiety. The results of their studies and others' did not clearly confirm this hypothesis. Winkel and Sarason found that female subjects performed better for low anxiety experimenters. Rosenthal's study demonstrated that medium

anxiety experimenters were the least effective influencers. In another study (Rosenthal, Persinger, Vikan-Kline, and Mulry, 1963) it was found that high anxiety experimenters were most influential on subjects' responses and that high anxiety subjects were most susceptible to being influenced. Finally, Rosenthal, Kohn, Greenfield, and Carota (1965) reported a study in which "less anxious" experimenters influenced subjects significantly better than "more anxious" experimenters. Thus, while anxiety of the experimenter appeared to be an important variable, its direction was unpredictable. The different results may reflect some difficulty in correctly identifying "high," "medium," and "low" anxiety experimenters. It is possible that experimenters affected their data in some unexamined way rather than by their anxiety level.

Need for social approval. Rosenthal conceptualized that experimenters who had a high need for social approval would be better influencers in order to be approved by the grand experimenter. This conceptualization was supported in his review of the literature (1967a). Need for social approval was not found to be related to the degree of experimenter influence, however, in a study by Rosenthal, Persinger, Vikan-Kline, and Mulry (1963).

Warmth. The warmth of experimenters seemed to be a significant variable. Reece and Whitman (1962) found that "warm" experimenters conditioned subjects significantly better ($p \leq .001$) than "cold" experimenters. Rosenthal, Kohn, Greenfield, and Carota (1965) reported that more friendly experimenters influenced subjects better. Authors of both studies felt that subjects were more likely to want to please warm, friendly experimenter.

Race of Experimenter

The influence of an experimenter's race has not been extensively studied. Two studies were reviewed which indicated that an experimenter's race alone does not affect data. Williams (1964) found that experimenter's race was significant only in interaction with a social distance (between interviewer and subject) variable, and when interview questions were potentially threatening. Katz, Robinson, Epps, and Waly (1964) administered a disguised verbal test of aggression to Negro high school males. When the test was neutrally described, Negro and white experimenters obtained the same results. When described as an intelligence test, Negro experimenters obtained significantly higher aggression scores than those elicited by white experimenters.

Cues Which Mediate Experimenter Influence

If experimenters do influence their research, the question must be asked, "How?" One factor analytic study (Rosenthal, Fode, Friedman, and Vikan-Kline, 1960) revealed that experimenters who read directions more slowly and used more hand gestures influenced their data the most. Thus, these are examples of "cues" which mediate some kind of experimenter influence. In more recent studies (Rosenthal and Fode, 1963a, 1963b; Rosenthal, Fode, Vikan-Kline, and Persinger, 1964; Friedman, Kurland, and Rosenthal, 1965; Rosenthal, Friedman, and Kurland, 1966; Friedman, 1967; and Rosenthal, 1967a) the following cues have been found significant in the transmission of experimenter influence: frequency of glances, rate and accuracy of reading directions, body activity, and touch. Rosenthal concluded that verbal

communication was the most significant mediator of experimenter influence.

That verbal cues were important mediators was confirmed by Sarason (1962). He found verbal reinforcement significantly more influential on subjects than a visual reinforcer (flashing light). Contrarily, Reece and Whitman (1962) reported a verbal conditioning study in which visual cues (body movement, smiling, glancing by experimenter) reinforced subjects more than verbal cues. Similarly, Carlsmith and Aronson (1963) found visual cues to be effective mediators of experimenter influence.

A majority of the authors of the studies reviewed in this section suggested that experimenters do influence their data. The results of these studies were too conflicting and inconsistent to warrant many definite conclusions regarding how experimenters influence their data.

The sex and race of experimenters seemed to be significant variables, especially when studied in interaction with other variables. The warmth of an experimenter appeared to have a consistent effect on data; that is, "warm" experimenters influenced their research more than experimenters perceived as being less warm.

The research on other experimenter personality characteristics and also on the prestige of the experimenter revealed little consistent evidence to support an assessment of their importance. Similarly, the effect of the experience of experimenter was discussed in an unresolved debate by Rosenthal, Harrington, and Ingraham. While visual and verbal cues were found to mediate experimenter influence, more research needs to be done to discover how experimenters influence

their data.

In several studies the grand experimenter deliberately gave experimenters an expectancy concerning the results of the study. There seemed to be two phenomena that created experimenter influence in these studies (Rosenthal, 1964): "expectancy effects," and "effects of early data returns."

Expectancy Effects

Rosenthal (1964) described "expectancy effects" when he reasoned that researchers usually studied variables in which they were interested and, therefore, often had certain "expectations" regarding that variable. If one's expectations in any way distorted the research data, experimenter influence existed.

To test the expectancy effect on data, Rosenthal, Mulry, Persinger, Vikan-Kline, and Grothe (1964b) presented a sequence of twenty photographs of people to a large norm group and asked them to rate each picture from -10 to +10 according to perceived failure (-10) or success (+10) exhibited by the faces. The mean rating was zero. In subsequent studies, experimenters were asked to administer the photographs to subjects and were told that the purpose was to develop an empathy test. Then experimenters were given differential expectancies concerning how their subjects would rate the photographs. That is, some experimenters were told to expect mean ratings of +5 (moderate success), and the other experimenters were led to expect -5 (moderate failure) mean ratings. Of course, subjects were assigned randomly and if significantly different mean ratings were obtained by +5 and -5 experimenters, a significant expectancy effect was

demonstrated. Using this basic design, Rosenthal and his students reported five studies (Rosenthal, Fode, Friedman, and Vikan-Kline, 1960; Rosenthal, Persinger, Vikan-Kline, and Mulry, 1963; Rosenthal, Mulry, Persinger, Vikan-Kline, and Grothe, 1964a, 1964b; and Friedman, Kurland, and Rosenthal, 1965) in which a significant experimenter expectancy effect was found. One study (Rosenthal, Fode, Vikan-Kline, and Persinger, 1964) revealed no significant effect.

Rosenthal and Halas (1962) and Cordaro and Ison (1963) found significant expectancy effects using planaria as subjects. Ingraham and Harrington (1966) reported no significant expectancy effect when experimenters were led to believe they had either "maze-bright" or "maze-dull" rats for subjects. With the same design, Rosenthal and Fode (1963a) had earlier found a significant difference: the supposedly "maze-bright" rats learned a discrimination task significantly faster.

Cooper, Eisenberg, Robert, and Dohrenwend (1967) gave opposite expectancy conditions to ten experimenters, each of whom had ten subjects. A significant expectancy effect was demonstrated. Using the photo-rating task described above, Friedman (1967) also found a significant expectancy effect.

Rosenthal and Jacobson (1966; 1968) examined the expectancy effect hypothesis in the classroom. They described their study at great length in Pygmalion in the Classroom (1968). The authors explained what they meant by an "interpersonal self-fulfilling prophecy" (1968, p. vii): " . . . how one person's expectation for another person's behavior can quite unwittingly become a more accurate prediction simply for its having been made." Specifically,

they wondered if a teacher's expectation of a pupil's ability somehow actually helped determine the pupil's ability. Working in an elementary school (Grades 1-6), Rosenthal and Jacobson administered a pretest (Flanagan Tests of General Ability) to all pupils in May of 1964, and told the teachers that it was the "Harvard Test of Inflected Acquisition." Then they randomly selected twenty percent of the pupils and told the teachers in September of 1964 that the test indicated that these pupils were about to take an "intellectual spurt" or were about to "bloom." Retesting occurred in January, 1965, the basic posttest was administered in May, 1965, and a follow-up posttest was administered in May, 1966. Rosenthal and Jacobson hypothesized that younger "bloomers" would show greater gains when compared to control pupils than older "bloomers." This hypothesis was confirmed in that significant results were found only at the first and second grade levels. Significant results were not obtained for an ability effect (fast, medium, slow tracks) or for a minority group status effect (Mexican versus American). Analysis did indicate a significant sex of pupil effect: boy "bloomers" spurted more than girl "bloomers." Rosenthal and Jacobson concluded that teachers apparently did communicate an expectation of performance to the "bloomers" which accounted for their gains on the intelligence test posttest. They conceptualized that the "quality of interaction" between teachers and "bloomers" probably made the difference.

In view of recent criticism (Thorndike, 1968; and Claiborn, 1969) of Rosenthal and Jacobson's study, one must continue to question the existence of teacher influence. Thorndike (1968, p. 711) concluded that the basic data were " . . . so untrustworthy that any

conclusions based upon them must be suspect." Claiborn (1969) replicated parts of the Rosenthal and Jacobson study and found no significant differences on the hypotheses tested in both studies. He discussed the "failure to replicate" and concluded that the question of teacher influence remained "equivocal." Taking the strength of Rosenthal's convictions and assuredness of his critics together, the writer was left rather confused. It seemed, however, that Rosenthal may be guilty of that which he has warned us about--observer influence.

Effects of Early Data Returns

"Early data return effect" referred to the tendency for experimenters to be influenced by the hypotheses suggested by data collected early in the life of an experiment. Rosenthal directed two ingenious studies of this effect. In the first study (Rosenthal, Persinger, Vikan-Kline, and Fode, 1963a) all experimenters were led to expect ± 5 mean photograph ratings. The researchers arranged for four of the experimenters to obtain "good" scores from their first two subjects by the use of coached accomplices and arranged for four experimenters to obtain "bad" scores by the same procedure. Four experimenters experienced only naive subjects who were not accomplices. The difference between the two experimental groups was significant and, as hypothesized, the control experimenters obtained a mean rating between the two experimental groups. Furthermore, there was some evidence of a sequence effect; that is, "good" data got better and "bad" data got worse. In a similar but more complex experiment (Rosenthal, Kohn, Greenfield, and Carota, 1965), the authors found a significant ($p \leq .05$) early data return effect. In this study, early data return

effect was strong enough to change an opposite, initial experimenter expectancy. The effect was strongest, however, when it confirmed an initial expectancy. The authors concluded that when experimenters experienced "good" data they became warmer and more friendly and exercised greater influence on their subjects.

C. Discussion of Previous Research

In this chapter studies of examiner influence on test scores were reviewed. There was only one study (Larrabee and Kleinsasser, 1967) reviewed in which the Wechsler Intelligence Scale for Children was used as an instrument.

Several studies concerning experimenter influence on experimental research were reviewed. The sex of the experimenter sometimes had an influence on experimental results when combined with other variables. Four studies were discussed in which the prestige of the experimenter was examined and that variable did not appear to have a consistent influence on research data. In a review of literature concerning the influence on research results of an experimenter's experience, it was concluded by this writer that experienced experimenters were less likely to influence their data. It was also found that certain personality characteristics (hostility, anxiety, need for social approval, and warmth) of experimenters sometimes influenced results of experiments.

If experimenter influence exists, it is important to understand how it is mediated. Studies were reported in which it was demonstrated that both verbal and non-verbal cues mediated experimenter influence.

Several studies were reviewed in which the grand experimenter

deliberately gave experimenters an expectancy regarding the experiment. In these studies an experimenter influence effect was consistently demonstrated.

This review of the literature revealed that experimenter influence has been extensively examined in experimental research. Rosenthal and Jacobson (1968) applied experimental research findings to a practical setting. They are vulnerable to criticism in regards to their research design and the interpretation of their data. The review of literature revealed weaknesses in research design of other studies as well. In many studies experimenters were given only one expectancy condition. This procedure was seen as a weakness because an expectancy "set" could easily develop which might account for the influencing effect. Much of the research involved male experimenters and female subjects of similar age. The sex of experimenter and sex of subject variables need to be crossed and younger subjects need to be tested. Most of the studies employed meaningless rating instruments which required little or no training to administer. This prevented an examination of the experimenter's experience or training as a variable in research on experimenter influence.

There seemed to be confusion concerning just how experimenter influence operated. Rosenthal (1967a) believed that experienced experimenters were more likely to influence their research because they were more ego-involved with their research and were "better biasers." That is, they were better at communicating their expectancy to subjects and better at reinforcing "correct" responses. Thus, for Rosenthal, experimenter influence increased with the experimenter's

experience and "snowballed" as an experiment progressed. Friedman (1967), one of Rosenthal's associates, agreed with this hypothesis.

This hypothesis has been continually criticized by Ingraham and Harrington (1966, 1967) in a running debate in Psychological Reports. They found (Ingraham and Harrington, 1966) that experimenters influenced less as an experiment progressed and as experimenters gained experience in the experimental task. They also found that training of experimenters reduced the experimenter influence effect and that experimenters who were given both expectancy conditions influenced their results less. They concluded (Harrington and Ingraham, 1967) that experimenter influence existed when an inexperienced and untrained experimenter began an experiment in which he was given an expectancy condition. As the experiment progressed, the experimenter responded more and more to factual cues presented by the subjects and less and less to the experimentally induced expectancy effect. Rosenthal (1967b, 1967c) retorted by manipulating Ingraham and Harrington's statistical methods in such a way that their research confirmed his hypothesis.

From this discussion it can be seen that the operation of experimenter influence is difficult to understand. For example, does the subject in an experimenter influence study actually perform better or does the experimenter just think the subject does better?

In Chapter III, the design of this study is presented. The sample, instrumentation, procedure, and the design and analysis are described. The research design was chosen to correct some of the weaknesses mentioned in the discussion of previous research.

CHAPTER III

DESIGN

Chapter III includes an explanation of the samples, instrumentation, procedure, and the design and analysis.

A. The Sample

Examiners

Eight trained Wechsler Intelligence Scale for Children (WISC) examiners participated in this study. They were volunteers from two Michigan State University individual testing classes (Education 866A) totalling approximately thirty-five people. Random selection of examiners would have been statistically desirable; however, due to insufficient funds, the writer could not make the research attractive enough to insure a large pool of examiners from which to randomly draw eight. Four men and five women volunteered to test eight examinees each for \$2.50 per test. A scheduling conflict eliminated one woman examiner which left four men and four women examiners.

Examinees

There were sixty-four examinees--thirty-two boys and thirty-two girls. Examinees were sixth, seventh, and eighth grade students at DeWitt Junior High School. DeWitt, Michigan is a middle-class suburb of Lansing, Michigan. The writer identified fifty-eight boys

and sixty-nine girls who had previously scored from ninety to one hundred-ten on the California Test of Mental Maturity--Short Form. From this population, thirty-two boys and thirty-two girls were randomly selected for the sample. The means and standard deviations obtained by examinees on the California Test of Mental Maturity are presented in Appendix A.

The examinees ranged in age from ten years, ten months to fourteen years, six months, with a median age of twelve years, seven and one-half months.

B. Instrumentation

Wechsler Intelligence Scale for Children

The Wechsler Intelligence Scale for Children (WISC) was used in this study for two reasons: first, a pool of trained WISC examiners was more accessible than Stanford-Binet or other individual intelligence test examiners; and second, although the WISC had high reliability, scoring was somewhat subjective. The nature of this study demanded that some degree of subjectivity in the scoring of protocols be permitted.

The WISC is an individual intelligence test for children between age five and age fifteen (Wechsler, 1949). There are five Verbal Scale subtests: Information, Comprehension, Arithmetic, Similarities, and Vocabulary. There are five Performance Scale subtests: Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding. In addition, there are two optional subtests: Digit Span and Mazes. Verbal, Performance, and Total I.Q.'s are computed by a standard score formula with mean set at 100 and the standard

deviation at fifteen. The WISC was standardized on a stratified random sampling of 2200 White American boys and girls. Published in 1949, the sampling was based on 1940 Census Bureau data.

Scoring of the WISC was made as objective as possible. However, Cronbach (1960, p. 194) pointed out that,

"The skill of the examiner may influence the score greatly. In some of the verbal tests, the examiner must make rather sensitive judgments as to the correctness of an answer since it may be necessary to request the subject to elaborate his meaning. Answers that seem wrong may be correct when the subject explains himself. Subjectivity in scoring borderline answers is also a potential problem."

Littell (1960) noted that the predictive validity of the WISC had not been demonstrated. Though Burstein (Buros, 1965) did not mention this limitation, no recent studies were reviewed in which an attempt was made to demonstrate the WISC's predictive validity. Concurrent validity has been successfully demonstrated. In various studies, the WISC correlates with the Stanford-Binet from .49 to .94 with a median correlation of about .80. Correlations between the WISC and the Wechsler-Bellevue ranged from .72 to .87. With the California Test of Mental Maturity, correlations ranged from .77 to .81. When correlated with achievement tests, coefficients ranged from .14 to .81 with a median correlation of .66.

WISC reliability is very high. Split-half reliability coefficients vary from .86 to .96 depending on the age level of examinees (Littell, 1960). Cronbach (1960, p. 198) felt that the Wechsler Performance Scale was the most reliable performance scale ever developed. Coefficients of internal consistency ranged from .59 to .91 with standard error of measurement ranging from 3.00 to 5.61. Only one coefficient of stability was reported (Littell, 1960) and it

was .77 over a four year period. Reliability coefficients were greatest from age ten and one-half to age thirteen and one-half. The subjects in this study were between the ages of ten and fourteen.

Fraser (Buros, 1959) felt that the WISC seemed most valid for normal range subjects. All the subjects in the present study were drawn from the "normal" range. In the Sixth Mental Measurement Yearbook (Buros, 1965), Burstein supported the WISC as having good reliability. He pointed out that much of the WISC research since 1960 concerned "psychopathological applications." That is, the WISC was used to measure the intelligence of special groups such as the retarded, the gifted, brain-damaged children, and the disadvantaged. A review of WISC literature since 1960 brought the writer to the same conclusion made earlier by Littell (1960): Studies need to be made concerning possible sources of score variation other than intelligence.

Rating Sheet

An expectancy influence was experimentally induced by giving test examiners a rating sheet for each examinee (see Appendix B). The rating sheet included the examinee's name and fictitious ratings of his California Test of Mental Maturity--Short Form I.Q. and school achievement. The fictitious ratings were marked on continua from "bottom one-fourth" to "top one-fourth." In each case a discrepancy between intelligence test score and school achievement was indicated. For example, an "above average" examinee was rated at the top one-fourth on the CTMM, and second one-fourth on school achievement. The last item on the rating sheet was a "predicted WISC score" which was

above average or below average depending on whether the examinee had previously been designated as "above average" or "below average." That is, examinees who were supposedly "above average" were given predicted WISC scores above one hundred. Examinees who were supposedly "below average" were assigned predicted WISC scores below one hundred. In this way it was hoped that examiners would have an "expectancy" regarding the outcome of an examinee's WISC score.

C. Procedure

After the thirty-two boys and thirty-two girls were selected, they were randomly assigned to one of the eight examiners so that each examiner had four boys and four girls. Then "above average" or "below average" designations were randomly assigned to examinees so that each examiner would test two "above average" boys, two "below average" boys, two "above average" girls, and two "below average" girls. Finally, the order in which examinees were tested was randomized for each examiner.

It was necessary to call seventy parents to secure approval for testing sixty-four subjects. When an originally selected subject could not participate, a replacement was randomly drawn from the remaining population. Testing was offered as a service to parents and confidentiality was assured. Neither the subjects nor their parents were told that the testing was part of a "study" or an "experiment." Test scores were interpreted to parents at a later date.

Examiners were given the following information before testing began:

"These students are being tested because there is a discrepancy between their I.Q. (CTMM) and school grades. Be sure to read the information for each student before testing because we are interested in knowing your opinion with regard to whether he has been over-achieving or under-achieving. When turning in test scores please let us know the student's I.Q. and any special circumstances concerning the Verbal and Performances scores."

Examiners were led to believe that the writer was merely an agent of the DeWitt Public Schools. In no case did an examiner indicate that he "guessed" that he was taking part in an experiment or was indeed a subject in an experiment.

All testing was completed within a one week period starting the day after examiners completed their WISC course. Most examiners tested four examinees in each of two sessions. Two examiners tested eight examinees in one day. Scheduling conflicts prevented a more uniform testing schedule which would have been desirable in terms of research design.

Examinees were scheduled at seventy-five or ninety minute intervals. Each examiner had a private office in which to administer the WISC. No more than three examiners were testing at one time and conversation among examiners was infrequent. It was feared that too much discussion among examiners might accidentally lead examiners to the realization that they were subjects in the research design. All tests were scored by the examiners and returned to the writer.

D. Design and Analysis

Analysis of the data was by a mixed model, four-way analysis of variance (Hays, pp. 439-447). A "mixed model" was used because the design had both "fixed" and "random" variables. A variable is

"fixed" when there are qualitatively distinct levels of the variable. A "random" variable is one in which the levels chosen represent only a sample of the population of possible levels. The design of the study (see Figure 1) consisted of three fixed variables (expectancy, sex of examinee, and sex of examiner), and two random variables (examiners nested in sex of examiner, and replications nested in all other variables). A variable is nested in a second variable when each level of the nested variable does not appear in all levels of the second variable. For example, from Figure 1 it can be seen that Examiner 2 can only be in one level of the sex of examiner variable because he is a male.

Sex of Examiner		Expectancy of Examiner by Examinee Sex			
		Male Examinee		Female Examinee	
		"Above Average" Expectancy	"Below Average" Expectancy	"Above Average" Expectancy	"Below Average" Expectancy
Male	1	R 1, R 2	R 1, R 2		
	2				
	3				
	4				
Female	5				
	6				
	7				
	8				

Fig. 1 Research Design

As seen in Figure 1, the sex of examiner variable is crossed with the sex of examinee and expectancy variables. Variables are "crossed" when each level of one variable occurs with each level of the other variable. From the design diagrammed in Figure 1 it can be seen that four male examiners and four female examiners each tested two "above average" male examinees, two "below average" males, two "above average" females, and two "below average" females. The five per cent level of confidence was arbitrarily chosen for significance tests.

Analysis of variance with fixed and random variables involves the usual assumptions of independence of observations, equality of variance, and normality of distributions. In less complex designs, at least, the violation of these assumptions in designs with equal numbers in the subgroups has been shown to have negligible effects on the significance tests. See, for example, Scheffe (1959, p. 354), Norton (1952), Young and Veldman (1963), and Boneau (1969).

Analysis of variance with fixed and random variables, as conducted here, involves the additional assumption (Hays, p. 463) that the degree of relationship, if any, among the different observations for the same fixed variable (expectancy, sex of examiner, and sex of examinees) are the same for all levels of that variable. There is no reason to believe this assumption is violated with the present data.

The use of analysis of variance in complex designs requires the assumption, accepted here and by most investigators, that the technique continues to be robust and that it is insensitive to violations of more basic assumptions in these applications.

The sources of variation and their degrees of freedom are presented in Table 1.

TABLE 1
SOURCES OF VARIATION AND
DEGREES OF FREEDOM

Sources of Variation	Degrees of Freedom
Sex of Examiner	1
Sex of Examinee	1
Expectancy	1
Examiner Nested in Sex of Examiner	6
Replications	32
Sex of Examiner X Sex of Examinee	1
Sex of Examiner X Expectancy	1
Sex of Examinee X Expectancy	1
Examiner X Expectancy Nested in Sex of Examiner	6
Examiner X Sex of Examinee Nested in Expectancy	6
Examiner X Expectancy X Sex of Exami- nee Nested in Sex of Examiner	6
Sex of Examiner X Expectancy X Sex of Examinee	1
Total	<u>63</u>

E. Summary

Chapter III contained an explanation of the samples, instrumentation, and analysis of the research design. The sample consisted of four male and four female Wechsler Intelligence Scale for Children (WISC) examiners who had just completed WISC training. Each examiner tested eight junior high school students (four boys and four girls) who were randomly selected from a population of students who were "average" in intelligence; that is, they scored from ninety to one hundred-ten on the California Test of Mental Maturity--Short Form, 1963 Revision. Examinees were randomly assigned to examiners.

The WISC is an individual intelligence test for children between ages five and fifteen. It has good reliability with split-half reliability coefficients ranging from .86 to .96. A rating sheet (see Appendix B) was used by the grand experimenter to transmit an expectancy condition to examiners. On each examinee's rating sheet a discrepancy between California Test of Mental Maturity--Short Form score and school achievement was fictitiously indicated, and a "predicted WISC score" was advanced.

Eight examinees were randomly assigned to each examiner and then they were randomly designated "above average" or "below average" so that each examiner tested two "above average" boys, two "below average" boys, two "above average" girls, and two "below average" girls. The order in which an examiner tested his examinees was also randomized. Parental approval for testing was secured and confidentiality was assured. Neither parents nor examiners were told that they were participating in a study or an experiment. All testing

was done in private offices during a one week period. Analysis of the data was by four-way analysis of variance.

In Chapter IV the hypotheses will be restated and the results will be presented.

CHAPTER IV

ANALYSIS OF RESULTS

In Chapter IV each hypothesis will be presented with a statement concerning whether the hypothesis in question will be rejected or not rejected. The significance of the obtained F-ratios will be determined by using Table IV in Hays (1964, pp. 677-688). Following this section, the results of the hypotheses will be discussed. Wechsler Intelligence Scale for Children (WISC) Total scores are presented in Appendix C.

A. Hypotheses and Results

Hypothesis 1

Female examiners will obtain a significantly higher mean Total WISC score than male examiners.

For a one-tailed F-test with one and six degrees of freedom, the required F was 5.99. Since the obtained F-ratio of 8.17 exceeded 5.99, the sex of examiner hypothesis was not rejected. The sex of examiner effect was significant ($p \leq .05$) in that female examiners obtained a higher mean Total WISC score than male examiners (see Table 2).

TABLE 2

SUMMARY OF MEAN TOTAL WISC SCORES, STANDARD DEVIATIONS,
AND F-RATIOS FOR MAIN EFFECTS

Main Effect	Means	Standard Deviations	F-Ratio
Sex of Examiner			
Males	97.53	8.75	
Females	103.12	8.06	8.17*
Sex of Examinee			
Males	101.66	9.03	
Females	99.00	9.07	9.30*
Expectancy			
"Above Average"	101.17	9.89	
"Below Average"	99.50	8.25	.84
Examiner			
1	97.75	7.49	
2	101.50	9.52	
3	94.00	8.60	
4	96.88	11.31	
5	104.75	7.83	
6	103.50	6.63	
7	104.62	9.98	
8	99.62	8.79	.92

*Sig. at .05 level of confidence

Hypothesis 2

There will be no significant difference between mean Total WISC scores achieved by boy and girl examinees.

For a two-tailed F-test with one and six degrees of freedom, the required F was 8.81. Since the obtained F-ratio was 9.30, the null hypothesis was rejected. The sex of examinee effect was significant ($p \leq .05$) in that male examinees achieved a higher mean Total WISC score than female examinees (see Table 2).

Hypothesis 3

The mean Total WISC score achieved by "above average" examinees will be significantly higher than the mean Total WISC score achieved by "below average" examinees ("above average" examinees were those whose predicted WISC score was above one hundred; "below average" examinees were those whose predicted WISC score was below one hundred).

For a one-tailed F-test with one and six degrees of freedom, the required F was 5.99. Since the obtained F-ratio was .84, the directional hypothesis was rejected. The expectancy effect was not significant (see Table 2).

Hypothesis 4

There will be no significant difference between mean Total WISC scores obtained by the eight examiners.

For a two-tailed F-test with six and thirty-two degrees of freedom, the required F was 2.87. Since the obtained F-ratio was .92, the null hypothesis was not rejected. The examiner effect was not significant (see Table 2).

Hypothesis 5

The mean Total WISC score obtained by opposite sex combinations of examiners and examinees will be significantly higher than the mean Total WISC score obtained by same sex combinations.

For a one-tailed F-test with one and thirty-two degrees of freedom, the required F was 4.17. Since the obtained F-ratio was .28, the directional hypothesis was rejected. The interaction effect between the sex of examiner and sex of examinee variables was not

significant (see Table 3).

TABLE 3

SUMMARY OF MEAN TOTAL WISC SCORES, STANDARD DEVIATIONS,
AND F-RATIOS FOR TWO-WAY INTERACTION EFFECTS

Interaction Effect		Means	Standard Deviations	F
Sex of Examiner X Sex of Examinee				
Male Examiner-Male Examinee		98.31	9.52	.28
Male Examiner-Female Examinee		96.75	9.28	
Female Examiner-Male Examinee		105.00	7.34	
Female Examiner-Female Examinee		100.75	12.76	
Sex of Examiner X Expectancy				
Male Examiner-"Above Average" Examinees		99.00	11.25	.39
Male Examiner-"Below Average" Examinees		96.06	6.86	
Female Examiner-"Above Average" Examinees		102.81	13.26	
Female Examiner-"Below Average" Examinees		102.94	8.27	
Sex of Examinee X Expectancy				
Male Examinee-"Above Average"		102.69	9.95	.01
Male Examinee-"Below Average"		100.62	8.19	
Female Examinee-"Above Average"		99.62	9.92	
Female Examinee-"Below Average"		98.38	8.41	

Hypothesis 6

There will be no significant interaction effect on mean Total WISC scores between the sex of the examiner and the extent to which the obtained Total WISC scores are in the direction suggested by the expectancy set.

For a two-tailed F-test with one and thirty-two degrees of freedom, the required F was 5.57. Since the obtained F was .39, the null hypothesis was not rejected. The interaction effect between the sex of examiner and expectancy variables was not significant (see Table 3).

Hypothesis 7

There will be no significant interaction effect on mean Total WISC scores between the sex of the examinee and the extent to which the obtained Total WISC scores are in the direction suggested by the expectancy set.

For a two-tailed F-test with one and six degrees of freedom, the required was 8.81. Since the obtained F-ratio was .01, the null hypothesis was not rejected. The interaction effect between the sex of examinee and expectancy variables was not significant (see Table 3).

Hypothesis 8

There will be no significant interaction effect among the sex of examiner, sex of examinee, and direction of expectancy variables on mean Total WISC scores.

For a two-tailed F-test with one and thirty-two degrees of freedom, the required F was 5.57. Since the obtained F was .01, the null hypothesis was not rejected. The three-way interaction effect between the sex of examiner, sex of examinee, and expectancy variables was not significant (see Table 4).

TABLE 4

SUMMARY OF MEAN TOTAL WISC SCORES, STANDARD DEVIATIONS,
AND F-RATIO FOR THREE-WAY INTERACTION EFFECT

Interaction Effect	Standard		F
	Means	Deviations	
Sex of Examiner X Sex of Examinee X Expectancy			
Male Examiner-Male Examinee-"Above Average"	100.12	12.12	
Male Examiner-Male Examinee-"Below Average"	96.50	6.30	
Male Examiner-Female Examinee-"Above Average"	97.88	11.02	
Male Examiner-Female Examinee-"Below Average"	95.62	7.79	
Female Examiner-Male Examinee-"Above Average"	105.25	7.09	
Female Examiner-Male Examinee-"Below Average"	107.75	8.07	
Female Examiner-Female Examinee-"Above Average"	100.28	17.19	
Female Examiner-Female Examinee-"Below Average"	101.12	8.59	.01

B. Discussion of Results

Female WISC examiners obtained significantly ($p \leq .05$) higher WISC scores from examinees than male examiners obtained. This result supported Cleutat's (1965) research with the Stanford-Binet. As Rosenthal (1966, p. 47) concluded, sex of experimenter seemed to be an "active" rather than a "passive" variable but not a very predictable one. The design of this study did not provide for determining whether significant effects were brought about by examinees or by examiners' perceptions of examinees. It was felt, however, that the WISC testing was a novel and, perhaps, an anxiety-producing event for most examinees. Junior high school students were asked to go to the nearly empty high school at a designated time and be tested by a complete stranger. A possible explanation is that examinees felt more comfortable in the presence of female examiners since most of their teachers were females. There were only three male teachers in Grades K through 8 of the school district. Three of four female

examiners obtained mean WISC scores higher than the highest mean WISC score obtained by any male examiner.

The significant ($p \leq .05$) sex of examinee effect was unexpected. Cicutat (1965) did not find a significant sex of examinee effect. It was conjectured that boys were more aggressive and competitive during the testing situation which accounted for their mean WISC score. No further attempt was made to explain this effect.

The central hypothesis of this study was that examinees who were allegedly "above average" would obtain a mean WISC score significantly higher than supposedly "below average" examinees. That is, did the grand experimenter's estimate of an examinee's WISC score create an expectancy on the part of the examiner which influenced the obtained WISC score? Though the results were in the desired direction, the expectancy effect did not approach significance. Several explanations for this failure to demonstrate examiner influence were made.

First, it must be considered possible that the WISC examiners did not receive and/or retain an expectancy regarding the outcome of an examinee's WISC score based on the grand experimenter's estimate of that WISC score. This possibility was believed very plausible. In criticizing Rosenthal's research, Barber and Silver (1968a, p. 25) outlined an "eight-step transmission process" involved in inducing experimenter influence.

"(a) The student experimenter attended to the expectancy communication from the principal investigator. (b) The experimenter comprehended the expectancy communication. (c) The experimenter retained the communication. (d) The experimenter (intentionally or unintentionally) attempted to transmit the expectancy to the subject. (e) The subject (consciously or unconsciously) attended to the expectancy communication from

the experimenter. (f) The subject (consciously or unconsciously) comprehended the experimenter's expectancy. (g) The subject (consciously or unconsciously) retained the experimenter's expectancy. (h) The subject (wittingly or unwittingly) acted upon (gave responses in harmony with) the experimenter's expectancy."

It was believed that WISC examiners "attended to the expectancy communication" and comprehended it. However, a break in the transmission process may have come when examiners failed to retain the expectancy communication. Therefore, little attempt was made to intentionally or unintentionally transmit the expectancy to examinees. The grand experimenter gave an expectancy for each examinee to the examiner at the very beginning of testing. It was believed that examiners listened and understood that each examinee would be either "above average" or "below average." Furthermore, the grand experimenter reinforced the expectancy condition for each examinee immediately preceding each test administration. However, once testing began, it was felt that examiners may have forgotten the expectancy regarding the examinee being tested. Examiners did not appear to be ego involved with the expectancy for each examinee. The possible failure to control and measure the expectancy condition must be regarded as a limitation of this study.

In addition to a breakdown in the transmission process, the failure to show a significant expectancy effect was possibly due to the nature of the experimental task and the experience of examiners. Ingraham and Harrington (1966, 1967), and Barber and Silver (1968a) suggested that expectancy effect was difficult to demonstrate in relatively structured tasks. Barber and Silver (1968a, p. 26) concluded, "Several studies in this area used relatively structured or

factual tasks, such as the Wechsler Adult Intelligence Scale, the Taylor Manifest Anxiety scale, and a number-estimation task; none of these studies showed an experimenter bias effect." Compared to Rosenthal's photo-rating task, the WISC was a very structured experimental task. Examiners did not have to respond to the ambiguous cues given them by the grand experimenter; rather, they very quickly responded to factual cues presented to them by examinees. Since these factual cues represented continuing information concerning an examinee's true intelligence, the "expected" WISC score for an examinee may have become increasingly less important and disregarded or forgotten.

Ingraham and Harrington (1966, 1967) also believed that experienced and well-trained experimenters did not allow their expectancies to influence their results as much as less experienced experimenters. Though the WISC examiners were inexperienced when compared to experienced WISC examiners, they were well-trained when compared to the experimenters employed in most of Rosenthal's research. Each examiner had just completed a five-week course in WISC-WAIS test administration. Presumably, they were trained to rigorously follow directions of administering the test and recording examinees' responses. They were trained not to intentionally influence or distort an examinee's responses. Examiners did not need to remember a "predicted" WISC score for an examinee because they would soon determine an "actual" WISC score for an examinee.

Barber and Silver (1968a, p. 26) postulated that an expectancy effect was easier to demonstrate when a subordinate-superordinate relationship existed between experimenters and the grand experimenter.

The WISC examiners in this study were responsible for eight test administrations but in no other way were they subordinate to the grand experimenter. Furthermore, the amount of compensation they received for testing was fixed and not dependent on the test results. In some studies experimenters received more compensation if their results confirmed the expectancy condition.

Finally, it was believed that the design of this study minimized the possibility of Type I error (rejecting null hypothesis when it should not be rejected) in regards to the expectancy effect. As mentioned above, the WISC was a relatively structured and factual experimental task, and examiners were relatively well-trained and experienced. In addition, each examiner tested both male and female examinees, and worked under both expectancy conditions which were randomly assigned. It was impossible for an examiner to think that all his examinees were "above average" or that they were all "below average."

Considering the failure to demonstrate a significant expectancy effect, the results of this study seemed to support the conclusions reached by Ingraham and Harrington (1966, 1967), and Barber and Silver (1968a, 1968b): experimenter influence was not as easy to demonstrate as Rosenthal claimed--especially in structured tasks using experienced experimenters.

The examiner effect was examined in the study to obtain evidence on Rosenthal's contention that this variable should always be tested in research, and also because adding this variable strengthened the research design. Though some variation existed among the mean Total WISC scores obtained by examiners, the F-ratio (.92) did not

approach significance. This insignificant result contraindicated the possibility that examiners influenced test scores in idiosyncratic ways. The result may also be taken as further evidence of the WISC's reliability.

Turning to a discussion of interaction effects, it was interesting to note that none of the four interaction effects tested was significant. The review of literature for this study abounded with significant interaction effects using marble sorting ability or photo ratings as the dependent variable. Again, the failure to demonstrate significant interaction effects was possibly due to the structured nature of the experimental task and the experience of the WISC examiners. Since no interaction effects were significant, no t-tests were computed between means.

There was considerable evidence in related literature to support the directional hypothesis that opposite sex combinations of examiners and examinees would obtain higher mean WISC scores than same sex combinations. This hypothesis was not supported in this study ($F = .28$). Though the female examiner-male examinee combination resulted in the highest mean Total WISC score (105.00), the male examiner-female examinee mean Total WISC score (96.75) was the lowest combination.

The interaction of the sex of examiner and expectancy variables was investigated to determine whether males or females influenced their data more. Rosenthal, Mulry, Persinger, Vikan-Kline, and Grothe (1964) found that males were better "biasers." With an F-ratio of .39, that conclusion was not supported in the present study. Examination of mean WISC scores (see Table 3) showed that male

examiners did obtain a mean WISC score for "above average" examinees that was almost three points above that obtained for "below average" examinees. There was practically no difference in mean WISC scores obtained for "above average" and "below average" examinees by female examiners. While the results were in the expected direction, the obtained F-ratio was far from significant and this discussion should not be interpreted as support for the hypothesis that males are better "biasers" than females. In Barber and Silver's (1968a, p. 25) terms, examiners apparently did not receive and/or retain an expectancy for examinees and therefore did not transmit an expectancy communication to examinees.

There was virtually no ($F = .01$) interaction effect between the sex of examinee and expectancy variables. Though a directional hypothesis was not advanced, it was postulated that female examinees' WISC performance would be more influenced by examiners than male examinees' WISC performance. Non-statistical examination of mean scores (see Table 3) revealed insignificant differences in the opposite direction. That is, the difference between "above average" and "below average" male examinees' mean WISC score was greater than the difference between "above average" and "below average" mean WISC scores for female examinees. As suggested earlier (p. 6), this result possibly contradicted previous research (Rosenthal, Mulry, Persinger, Vikan-Kline, and Grothe, 1964a; and Stevenson and Allen, 1964) because of the greater age difference between examiners and examinees in this study. Specifically, male college experimenters may have more cues by which to influence female college subjects than a junior high school girl. Of course one could counter with

the hypothesis that the greater age difference between examiner and examinee in this study could increase the subordinate-superordinate relationship which should increase examiner influence. A more plausible explanation for the insignificant interaction effect was believed to be that the WISC was a structured task and that the expectancy effect demonstrated in Hypothesis 3 was just too weak. Apparently the expectancy communication was not received by the examinees of this study.

The three-way interaction effect was investigated to determine if the three variables (sex of examiner, sex of examinee, and expectancy) were working in some complex way. No significant interaction effect ($F = .01$) was found. The hypothesis was advanced for exploratory reasons and no attempt was made to explain the insignificant effect.

C. Summary

In Chapter IV the statistical hypotheses were restated and a decision was made to reject or not reject each hypothesis. Then the results of each hypothesis were discussed in reference to previous research and possible explanations for the results.

Both the sex of examiner and sex of examinee effects were significant ($p \leq .05$). Female examiners obtained higher mean WISC scores and male examinees achieved higher mean WISC scores. While the significant sex of examiner effect confirmed previous research (Cleutat, 1965), the significant sex of examinee effect was unexpected. It was conceptualized that male examinees were less anxious and more competitive when placed in a novel situation. The

interaction effect between sex of examiner and sex of examinee was not significant. The main sex effects seemed to over-shadow the interaction effect.

The major finding in this study was that the expectancy effect was not significant. Two principle explanations were offered. First, the eight-step transmission process delineated by Barber and Silver (1968a, p. 25) broke down in this study. That is, it was feared that WISC examiners did not retain the expectancy condition given to them by the grand experimenter and, therefore, examiners' expectancy was neither transmitted to nor received by examinees. Secondly, it was felt that the WISC represented a structured and factual experimental task, and that the WISC examiners were relatively well-trained and experienced. Thus, the results of the study seemed to confirm Ingraham and Harrington's (1966, 1967) and Barber and Silver's (1968a, 1968b) criticism of Rosenthal's research: experimenter influence has not been consistently demonstrated when experienced experimenters are administering structured tasks.

The fact that no significant examiner effect was demonstrated in this study added to the argument that examiner influence was not a great problem in intelligence testing, at least not in this study.

The expectancy variable was examined in interaction with both the sex of examiner and sex of examinee variables and neither interaction effect was significant. Thus, the results of this study were not construed as evidence that male experimenters were better influencers or that female examinees were more easily influenced. These two insignificant interaction effects represented further evidence of a lack of an expectancy effect in situations having this degree of

of examination structure and examiner experience. Analysis of the three-way interaction effect among the sex of examiner, sex of examinee, and expectancy variables revealed no significant complex interaction.

Chapter V will include a summary of this study and conclusions made from it.

CHAPTER V

SUMMARY, CONCLUSIONS, AND IMPLICATION FOR FURTHER RESEARCH

A. Summary

The present study was made to examine the claim of Robert Rosenthal (1967a) and others that experimenters frequently influenced (intentionally or unintentionally) the results of their research. Cloutat (1965), Cohen (1965), and Larrabee and Kleinsasser (1967) investigated the possibility that examiners of individual intelligence tests affected the scores of examinees.

The purpose of this study was to provide a well-designed field testing of experimental research results. Specifically, the writer sought to determine if Wechsler Intelligence Scale for Children (WISC) examiners influenced their examinees' scores when the examiners were given an expectancy or estimate of each examinee's score.

Eight hypotheses were advanced:

Hypothesis 1

Female examiners will obtain a higher mean intelligence test score than male examiners.

Hypothesis 2

There will be no difference between mean intelligence test scores achieved by boy and girl examinees.

Hypothesis 3

Examinees who are expected by examiners to be above average will achieve a higher mean intelligence test score than examinees who are expected to be below average.

Hypothesis 4

There will be no difference between mean intelligence test scores obtained by the examiners.

Hypothesis 5

Opposite sex combinations of examiners and examinees will obtain higher mean intelligence test scores than same sex combinations.

Hypothesis 6

There will be no interaction effect between the sex of the examiner and the extent to which the obtained scores are in the direction suggested by the expectancy set.

Hypothesis 7

There will be no interaction effect between the sex of the examinee and the extent to which the obtained scores are in the direction suggested by the expectancy set.

Hypothesis 8

There will be no interaction effect among the sex of examiner, sex of examinee, and direction of expectancy variables.

Chapter II contained a review of recent (1960 to present) literature concerning examiner influence on individual intelligence

test scores and on projective test score; research on experimenter influence; and a discussion of previous research.

Examiner influence on individual intelligence test scores and on projective test scores was investigated in a limited number of research studies. In these studies the examiner's personality (warmth and rapport) or physical characteristics (sex and race) had an effect on the dependent variable. Specifically, opposite sex combinations and same race combinations produced the highest scores.

A review of studies of experimenter influence revealed that experimenters frequently appear to influence the results of their studies. However, the results were too conflicting and inconsistent to warrant many definite conclusions regarding how experimenters influence their data. The sex and race of experimenters seemed to be significant variables, especially when studied in interaction with other variables. The warmth of an experimenter appeared to have a consistent effect on data; that is, "warm" experimenters influenced their research more than experimenters perceived as being less warm.

The research on other experimenter personality characteristics and also on the prestige of the experimenter revealed little consistent evidence to support an assessment of their importance. Similarly, the effect of the experience of experimenter was discussed in an unresolved debate by Rosenthal, Harrington, and Ingraham. While visual and verbal cues were found to mediate experimenter influence, more research needs to be done to discover how experimenters influence their data.

The review of studies in which a grand experimenter deliberately gave experimenters an expectancy concerning the results of the study consistently revealed a significant expectancy effect.

Rosenthal and Jacobson's (1968) Oak School experiment was discussed as a study in which the authors attempted to apply experimental results to a practical setting. Rosenthal and Jacobson concluded that the teachers in the experiment had communicated an expectation of ability to "bloomers" which accounted for their intellectual "spurt."

In a discussion of previous research some weaknesses in research design were mentioned. In most studies experimenters were given only one expectancy condition. This procedure was seen as a weakness because an expectancy "set" could develop which might account for the influence effect. Much of the research involved male experimenters and female subjects of similar age. The sex of experimenter and sex of subject variables need to be crossed and younger subjects need to be tested. Most of the studies employed meaningless rating instruments which required little or no training to administer. This prevented an examination of the experimenter's experience or training as a variable in research on experimenter influence.

There seemed to be confusion concerning just how experimenter influence operated. Rosenthal (1967a) believed that experienced experimenters were likely to influence their research because they were more ego-involved with their research and were "better biasers." That is, they were better at communicating their expectancy to subjects and better at reinforcing "correct" responses. Thus, for Rosenthal, experimenter influence increased with the experimenter's experience and "snowballed" as an experiment progressed. Friedman

(1967), one of Rosenthal's associates, agreed with this hypothesis.

This hypothesis was sharply criticized by Ingraham and Harrington (1966, 1967) and by Barber and Silver (1968a, 1968b). These authors felt that experimenters who were experienced and well-trained relied more on factual cues given them during an experiment and less on ambiguous cues given them by grand experimenters.

Chapter III included a discussion of the samples, instrumentation, procedure, and the design and analysis of the study.

The sample consisted of four male and four female Wechsler Intelligence Scale for Children (WISC) volunteer examiners who had just completed WISC training. They were paid \$2.50 per test. Each examiner tested eight junior high school students (four boys and four girls) who were randomly selected from a population of students who were "average" in intelligence (that is, they scored from ninety to one hundred-ten on the California Test of Mental Maturity--Short Form, 1963 Revision). Examinees were randomly assigned to examiners.

The WISC is an individual intelligence test for children between ages five and fifteen. It was selected for use in this study because WISC examiners were available and because the WISC had good reliability (split-half reliability coefficients ranged from .86 to .96). A rating sheet (see Appendix B) was used by the grand experimenter to transmit an expectancy condition to examiners. On each examinee's rating sheet a discrepancy between California Test of Mental Maturity--Short Form score and school achievement was fictitiously indicated, and a "predicted WISC score" was advanced.

After eight examinees were randomly assigned to an examiner,

they were randomly designated "above average" or "below average" so that each examiner tested two "above average" boys, two "below average" boys, two "above average" girls, and two "below average" girls. The order in which an examiner tested his examinees was also randomized.

Parental approval for testing was secured and confidentiality was assured. Neither parents nor examiners were told that the testing was part of a research study or experiment. In no case did an examiner indicate that he "guessed" that he was taking part in an experiment. All testing was done in private offices at the high school during a one week period.

Results were analyzed by four-way analysis of variance. The variables were sex of examiner, sex of examinee, expectancy ("above average" or "below average"), examiners nested in sex of examiner, and replications nested in all other variables. The five per cent level of significance was arbitrarily chosen.

In Chapter IV the eight hypotheses were restated in statistical form and rejected or not rejected. Both the sex of examiner and sex of examinee effects were significant ($p \leq .05$). As seen in Table 5 on the next page, female examiners obtained higher mean WISC scores and male examinees achieved higher mean WISC scores. While the significant sex of examiner effect confirmed previous research (Cieutat, 1965), the significant sex of examinee effect was unexpected. It was conceptualized that male examinees were less anxious and more competitive when placed in a novel situation. The interaction effect between sex of examiner and sex of examinee was not significant. The main sex effects seemed to over-shadow the

interaction effect.

TABLE 5
SUMMARY OF MEANS, STANDARD DEVIATIONS, AND F-RATIOS
FOR SIGNIFICANT EFFECTS

Effect	Means	Standard Deviations	F-Ratio
Sex of Examiner			
Male	97.53	8.75	
Female	103.12	8.06	8.17*
Sex of Examinee			
Male	101.66	9.03	
Female	99.00	9.07	9.30*

*Sig. at .05 level of confidence

The major finding in this study was that the expectancy effect was not significant. Two principle explanations were offered. First, the eight-step transmission process delineated by Barber and Silver (1968a, p. 25) broke down in this study. That is, it was feared that WISC examiners did not retain the expectancy condition given to them by the grand experimenter and, therefore, examiners' expectancy was neither transmitted to nor received by examinees. Secondly, it was felt that the WISC examiners were relatively well-trained and experienced. Thus, the results of the study seemed to confirm Ingraham and Harrington's (1966, 1967) and Barber and Silver's (1968a, 1968b) criticism of Rosenthal's research: experimenter influence has not been consistently demonstrated when experienced experimenters are administering structured tasks.

The fact that no significant examiner effect was demonstrated in this study added to the argument that examiner influence was not

a great problem in intelligence testing, at least not in this study.

The expectancy variable was examined in interaction with both the sex of examiner and sex of examinee variables and neither interaction was significant. Thus, the results of this study were not construed as evidence that male experimenters were better "influencers" or that female examinees were more easily influenced. These two insignificant interaction effects represented further evidence that the transmission of an expectancy condition did not take place, and that the WISC was a structured task administered by experienced examiners. Analysis of the three-way interaction effect among the sex of examiner, sex of examinee, and expectancy variables revealed no significant complex interaction.

B. Conclusions

The significant ($p \leq .05$) sex of examiner effect supported the feeling that the sex of experimenter or examiner was an active, though usually unpredictable variable. The mean Total WISC score obtained by the female examiners in this study was 5.59 IQ points higher than the mean Total WISC score obtained by male examiners. If important and somewhat irreversible decisions were going to be made for an individual on the basis of a WISC score, it would seem wise to have that individual tested by both a male and a female examiner. The added expense, effort, and time would be justified in relation to the importance of the decision to be made.

The fact that a significant expectancy effect was not demonstrated in this study confirmed Barber and Silver's (1968a, p. 26) conclusion that expectancy influence was not present when the experimental task was structured and factual. The insignificant

expectancy effect also supported Ingraham and Harrington's (1966, 1967) contention that experienced experimenters were not as likely to influence their results as inexperienced experimenters. The WISC examiners in this study did not retain the expectancy for an examinee when they started the testing situation. Instead, it was hypothesized that they increasingly responded to the factual stimuli presented them by examinees.

It was concluded that a well-designed study would minimize the probability of a significant expectancy effect. Specifically, it was considered important to design a study employing both male and female examiners, and both male and female examinees. The sex of experimenter and sex of subject variables were infrequently crossed in previous research. In agreement with Ingraham and Harrington (1966) was the feeling that it was necessary to give examiners both expectancy conditions in random order to prevent development of an expectancy "set."

It was also felt that many of the studies which demonstrated a significant expectancy effect lacked rigorous statistical methodology. Barber and Silver (1968a, p. 24) stated that, "(1) the variables to be studied and the statistics to be used should be specified in advance; (2) the level of significance should be stated in advance; (3) the data should be analyzed by some 'overall' test such as multivariate analysis of variance; and (4) conclusions should not be made from the results of post hoc tests performed upon the data after an overall test has failed to reject the null hypothesis." In the present study results were analyzed by an overall test

(four-way analysis of variance), and no "post hoc tests" were performed when the overall test failed to reject the null hypotheses.

Rosenthal (1967a) and other experimenters apparently assumed that if experimenter influence could be demonstrated in laboratory tasks, the results could be generalized or applied to practical and more meaningful situations such as individual intelligence testing. On the basis of this study and criticisms leveled by others (Barber and Silver, 1968a, 1968b; Barber, et.al., 1969; and Claiborn, 1969), this assumption appeared unwarranted. Rosenthal (1967a) often maximized the probability of obtaining a significant expectancy effect by (1) designing ambiguous experimental tasks; (2) using opposite sex combinations of experimenters and subjects; (3) making experimenters subordinate to the grand experimenter; (4) giving experimenters only one expectancy condition; or even by (5) paying experimenters more if their results were in the desired direction. Under a combination of these conditions, it was not surprising that a significant expectancy effect was found. However, in practical settings (such as individual intelligence testing) it was believed that these conditions usually did not exist. The generalizability of Rosenthal's experimental research was considered questionable and was not supported in this study. In fact, in light of the well-reasoned criticisms of Rosenthal's basic experimenter influence research (Ingraham and Harrington, 1966, 1967; Barber and Silver, 1968a, 1968b; and Barber, et.al., 1969), one must consider the possibility that observer influence and expectancy effect simply may not exist.

C. Implications for Further Research

Any replication of the present study should include a more adequate method of communicating the expectancy condition to WISC examiners and checking its presence during actual testing. Examiners ego-involvement in the expectancy condition given them must be increased and sustained. Controls should be included which derive from Barber and Silver's (1968a, p. 25) "eight-step transmission process." Before the existence of examiner influence, if any, can be demonstrated, future researchers must be certain that the expectancy communication was "attended to," "comprehended," and "retained" by test examiners and "transmitted" to examinees.

More importantly, further research concerning the sex of the examiner as a variable influencing individual intelligence test scores should be conducted. A research design employing male and female examiners testing the same examinees with alternate forms of the same test might be profitable.

From the results of their study, Rosenthal and Jacobson (1968) implied that more research should be conducted in teacher training programs to explore how teachers' expectations of pupils' ability affects actual pupil performance. The results of this study, however, did not imply a great need for further research on the problem of an expectancy effect on individual intelligence testing. It was not claimed that this study settled the question "once and for all;" however, the insignificant expectancy effect did support the conclusions of other researchers (Ingraham and Harrington, 1966, 1967; and Barber and Silver, 1968a, 1968b). Experimenter influence was not a

problem in structured, factual tasks administered by experienced and well-trained experimenters.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Barber, T.X. Invalid arguments, postmortem analyses, and the experimenter bias effect. Journal of Consulting and Clinical Psychology, 1969, 33, 11-14.
- Barber, T.X., Forgione, A., Chaves, J.F., Calverley, D.S., and McPeake, J.D. & Bowen, B. Five attempts to replicate the experimenter bias effect. Journal of Consulting and Clinical Psychology, 1969, 33, 1-6.
- Barber, T.X., & Silver, M.J. Fact, fiction, and the experimenter bias effect. Psychological Bulletin Monograph, 1968a, 70, 1-29.
- Barber, T.X., & Silver, M.J. Pitfalls in data analysis and interpretation: a reply to Rosenthal. Psychological Bulletin Monograph, 1968b, 70, 48-62.
- Bass, B.M. Measures of average influence and change in agreement of rankings by a group of judges. Sociometry, 1960, 23, 195-202.
- Blaufarb, H. The relation of experimenter status and achievement imagery to the conditioning of verbal behavior. Unpublished Ph.D. dissertation, Univ. of Illinois, 1960.
- Boneau, C.A. The effects of violations of assumptions underlying the t-test. Psychological Bulletin, 1969, 57, 49-64.
- Buros, O.K. (ed.). The Fifth Mental Measurements Yearbook. Highland Park, N.J.: The Gryphon Press, 1959.
- Buros, O.K. (ed.). The Sixth Mental Measurements Yearbook. Highland Park, N.J.: The Gryphon Press, 1965.
- Brogden, W.J. The experimenter as a factor in animal conditioning. Psychological Reports, 1962, 11, 239-242.
- Carlsmith, J.M., & Aronson, E. Some hedonic consequences of the confirmation and disconfirmation of expectancies. Journal of Abnormal and Social Psychology, 1963, 66, 151-156.
- Cleutat, V.J. Examiner differences with individual intelligence tests. Perceptual and Motor Skills, 1965, 20, 1317-1318.
- Claiborn, W.L. Expectancy effects in the classroom: a failure to replicate. Journal of Educational Psychology, 1969, 60, 377-383.

- Cohen, E. Examiner differences with individual intelligence tests. Perceptual and Motor Skills, 1965, 20, 1324.
- Cooper, J., Eisenberg, L., Robert, J., & Dohrenwend, B. The effect of experimenter expectancy and preparatory effort on belief in the probable occurrence of future events. Journal of Social Psychology, 1967, 71, 221-226.
- Cordaro, L., & Ison, J.R. Psychology of the scientist: X. Observer bias in classical conditioning of the planarian. Psychological Reports, 1963, 13, 787-789.
- Cronbach, L.J., Essentials of Psychological Testing. Second edition. New York: Harper and Brothers, 1960.
- Exner, J.E. Jr., Variations in WISC performances as influenced by differences in pretest rapport. Journal of Genetic Psychology, 1966, 74, 299-306.
- Friedman, N. The Social Nature of Psychological Research. New York: Basic Books, Inc., 1967.
- Friedman, N., Kurland, D., & Rosenthal, R. Experimenter behavior as an unintended determinant of experimental results. Journal of Projective Techniques and Personality Assessment, 1965, 29, 479-490.
- Griffith, R.M. Rorschach water percepts: a study in conflicting results. American Psychologist, 1961, 16, 307-311.
- Harrington, G.M., & Ingraham, L.H., Psychology of the scientist: XXV. Experimenter bias and tails of Pascal. Psychological Reports, 1967, 21, 513-516.
- Hays, W.L. Statistics for Psychologists. New York: Holt, Rinehart, & Winston, 1964.
- Hill, K.T., & Stevenson, H.W. The effects of social reinforcement vs. non-reinforcement and sex of E on the performance of adolescent girls. Journal of Personality, 1965, 33, 30-36.
- Ingraham, L.H., & Harrington, G.M. Psychology of the scientist: XVI. Experience of E as a variable in reducing experimenter bias. Psychological Reports, 1966, 19, 455-461.
- Jones, E.L. The courtesy bias in South-East Asian surveys. International Social Science Journal, 1963, 15, 70-76.
- Katkin, E.S., Risk, R.T., & Spielberger, C.D. The effects of experimenter status and subject awareness on verbal conditioning. Journal of Experimental Research in Personality, 1966, 1, 153-160.

- Katz, I., Robinson, J., Epps, E., & Waly, P. Race of experimenter and instructions in the expression of hostility by Negro boys. Journal of Social Issues, 1964, 20, 54-60.
- Kintz, B.L., Delprato, D.J., Mettee, D.R., Persons, C.E., & Schappe, R.H. The experimenter effect. Psychological Bulletin, 1965, 63, 223-232.
- Kish, L. Studies of interviewer variance for attitudinal variables. Journal of the American Statistical Association, 1962, 57, 92-115.
- Larrabee, L.L., & Kleinsasser, L.D. The effect of experimenter bias on WISC performance. Unpublished paper, St. Louis, Mo: Psychological Associates, 1967.
- Levy, L.H. Reflections on replications and the experimenter bias effect. Journal of Consulting and Clinical Psychology, 1969 33, 11-14.
- Littell, W.A. The Wechsler Scale for Children: review of a decade of research. Psychological Bulletin, 1960, 57, 132-156.
- Masling, J. The influence of situational and interpersonal variables in projective testing. Psychological Bulletin, 1960, 57, 65-85.
- McFall, R.M. Unintentional communication: the effect of congruence and incongruence between subject and experimenter constructions. Unpublished Ph.D. dissertation, Ohio State Univ., 1966.
- McGuigan, F.J. The experimenter: a neglected stimulus object. Psychological Bulletin, 1963, 60, 421-428.
- Miller, M.E., & Solkoff, N. Effects of mode of response and sex of E upon recognition thresholds of taboo words. Perceptual and Motor Skills, 1965, 20, 573-578.
- Mulry, R.C. The effects of the experimenter's perception of his own performance on subject performance in a pursuit rotor task. Unpublished master's thesis, Univ. of North Dakota, 1962.
- Norton, D.W. An empirical investigation of the effects of non-normality and heterogeneity upon the F-test of analysis of variance. Ph.D. thesis, State Univ. of Iowa, 1952.
- Ogawa, J., & Oakes, W.F. Sex of experimenter and manifest anxiety as related to verbal conditioning. Journal of Personality, 1965, 33, 553-569.
- Prince, A.I. Relative prestige and the verbal conditioning of children. Paper read at the Seventieth Annual Convention of the American Psychological Association, St. Louis, Missouri, 1962.

- Reece, M.M., & Whitman, R.N. Expressive movements, warmth, and verbal reinforcement, Journal of Abnormal and Social Psychology, 1962, 64, 234-236.
- Rosenthal, R. Covert communication in the psychological experiment. Psychological Bulletin, 1967a, 67, 356-367.
- Rosenthal, R. Experimenter attributes as determinants of subjects' responses. Journal of Projective Techniques and Personality Assessment, 1963a, 27, 324-331.
- Rosenthal, R. Experimenter Effects in Behavioral Research. New York: Appleton-Century-Crofts, 1966.
- Rosenthal, R. Experimenter expectancy and the reassuring nature of the null hypothesis decision procedures. Psychological Bulletin Monograph, 1968, 70, 30-47.
- Rosenthal, R. Experimenter modeling effects as determinants of subjects' responses. Journal of Projective Techniques and Personality Assessment, 1963b, 27, 467-471.
- Rosenthal, R. Experimenter outcome-orientation and the results of the psychological experiment. Psychological Bulletin, 1964, 61, 405-412.
- Rosenthal, R. On not so replicated experiments and not so null results. Journal of Consulting and Clinical Psychology, 1969, 33, 7-10.
- Rosenthal, R. Psychology of the scientist: XXIII. Experimenter expectancy, experimenter experience, and Pascal's wager. Psychological Reports, 1967b, 20, 619-622.
- Rosenthal, R. Psychology of the scientist: XXVI. Experimenter expectancy, one tail of Pascal, and the distribution of three tails. Psychological Reports, 1967c, 21, 517-520.
- Rosenthal, R., & Fode, K.L. The effect of experimenter bias on the performance of the albino rat. Behavioral Science, 1963a, 8, 183-189.
- Rosenthal, R., & Fode, K.L. Psychology of the scientist: V. Three experiments in experimenter bias. Psychological Reports, 1963b, 12, 491-511.
- Rosenthal, R., Fode, K.L., Friedman, C.J., & Vikan-Kline, L.L. Subjects' perception of their experimenter under conditions of experimenter bias. Perceptual and Motor Skills, 1960, 11, 325-331.
- Rosenthal, R., Fode, K.L., Vikan-Kline, L.L., & Persinger, G.W. Verbal conditioning: mediator of experimenter expectancy effects. Psychological Reports, 1964, 14, 71-74.

- Rosenthal, R., Friedman, C.J., Johnson, C.J., Fode, K.L., Schill, T., White, R.C., & Vikan-Kline, L.L. Variables affecting experimenter bias in a group situation. Genetic Psychology Monographs, 1964, 70, 271-296.
- Rosenthal, R., Friedman, N., & Kurland, D. Instruction-reading behavior of the experimenter as an unintended determinant of experimental results. Journal of Experimental Research in Personality, 1966, 1, 221-226.
- Rosenthal, R., & Halas, E.S. Experimenter effect in the study of invertebrate behavior. Psychological Reports, 1962, 11, 251-256.
- Rosenthal, R., & Jacobson, L. Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development. New York: Holt, Rinehart, and Winston, 1968.
- Rosenthal, R., & Jacobson, L. Teachers' expectancies: determinants of pupils' IQ. gains. Psychological Reports, 1966, 19, 115-118.
- Rosenthal, R., Kohn, P., Greenfield, P.M., & Carota, N. Data desirability, experimenter expectancy, and the results of psychological research. Journal of Personality and Social Psychology, 1966, 3, 20-27.
- Rosenthal, R., Kohn, P., Greenfield, P.M., & Carota, N. Psychology of the scientist: XIV. Experimenters' hypothesis-confirmation and mood as determinants of experimental results. Perceptual and Motor Skills, 1965, 20, 1237-1252.
- Rosenthal, R., Mulry, R.C., Persinger, G.W., Vikan-Kline, L.L., & Grothe, M. Changes in experimental hypotheses as determinants of experimental results. Journal of Projective Techniques and Personality Assessment, 1964a, 28, 465-469.
- Rosenthal, R., Mulry, R.C., Persinger, G.W., Vikan-Kline, L.L., & Grothe, M. Emphasis on experimental procedure, sex of subjects and the biasing effects of experimental hypotheses. Journal of Projective Techniques and Personality Assessment, 1964b, 28, 470-473.
- Rosenthal, R., & Persinger, G.W. Let's pretend: subjects' perception of imaginary experimenters. Perceptual and Motor Skills, 1962, 14, 407-409.
- Rosenthal, R., Persinger, G.W., & Fode, K.L. Experimenter bias, anxiety, and social desirability. Perceptual and Motor Skills, 1962, 15, 73-74.
- Rosenthal, R., Persinger, G.W., Vikan-Kline, L.L., & Fode, K.L. The effect of early data returns on data subsequently obtained by outcome-biased experimenters. Sociometry, 1963a, 26, 487-498.

- Rosenthal, R., Persinger, G.W., Vikan-Kline, L.L., & Fode, K.L. The effect of experimenter outcome-bias and subject set on awareness in verbal conditioning experiments. Journal of Verbal Learning and Verbal Behavior, 1963b, 2, 275-283.
- Rosenthal, R., Persinger, G.W., Vikan-Kline, L.L., & Mulry, R.C. The role of the research assistant in the mediation of experimenter bias. Journal of Personality, 1963, 31, 313-335.
- Sarason, I.G. Individual differences, situational variables, and personality research. Journal of Abnormal and Social Psychology, 1962, 65, 376-380.
- Sarason, I.G., & Minard, J. Interrelationships among subject, experimenter, and situational variables. Journal of Abnormal and Social Psychology, 1963, 67, 87-91.
- Scheffe, H. The Analysis of Variance. New York: Wiley, 1959.
- Simmons, W.L., & Christy, E.G. Verbal reinforcement of a TAT theme. Journal of Projective Techniques and Personality Assessment, 1962, 26, 337-341.
- Stevenson, H.W. Social reinforcement with children as a function of CA, sex of E, and sex of S. Journal of Abnormal and Social Psychology, 1961, 63, 147-154.
- Stevenson, H.W., & Allen, S. Adult performance as a function of sex of experimenter and sex of subject. Journal of Abnormal and Social Psychology, 1964, 68, 214-216.
- Stevenson, H.W., & Allen, S. Variables associated with adults' effectiveness as reinforcing agents, Journal of Personality, 1967, 35, 246-264.
- Suchman, E.A. An analysis of "bias" in survey research. Public Opinion Quarterly, 1962, 26, 102-111.
- Thorndike, R.L. Review of Rosenthal, R. & Jacobson, L. Pygmalion in the Classroom. American Educational Research Journal, 1968, 5, 708-711.
- Turner, G.C., & Coleman, J.C. Examiner influence on Thematic Apperception Test responses. Journal of Projective Techniques and Personality Assessment, 1962, 26, 478-486.
- Wartenberg-Ekren, U. The effect of experimenter knowledge of a subject's scholastic standing on the performance of a reasoning task. Unpublished master's thesis, Marquette Univ., 1962.
- Wechsler, D. WISC Manual. New York: The Psychological Corporation, 1949.

- Wilkie, C.H. A study of distortions in recording interviews. Social Work, 1963, 8, 31-36.
- Williams, J.A. Interviewer-respondent interaction: a study of bias in the information interview. Sociometry, 1964, 27, 338-352.
- Winkel, G.H., & Sarason, I.G. Subject, experimenter, and situational variables in research on anxiety. Journal of Abnormal and Social Psychology, 1964, 68, 601-608.
- Young, R.K., & Veldman, D.J. Heterogeneity and skewness in analysis of variance. Perceptual and Motor Skills, 1963, 16, 588.

APPENDICES

APPENDIX A

APPENDIX A

TABLE 6
CTMM MEANS AND STANDARD DEVIATIONS
FOR ALL SUBGROUPS

Effect	Means	Standard Deviations
Sex of Examiner		
Males	101.78	5.81
Females	102.00	5.61
Sex of Examinee		
Males	100.97	6.03
Females	102.81	5.21
Expectancy		
"Above Average"	102.88	5.29
"Below Average"	100.91	5.94
Examiner		
1	103.38	6.19
2	99.62	3.53
3	101.25	5.96
4	102.88	7.22
5	103.75	3.86
6	101.75	5.42
7	101.63	5.50
8	100.88	7.72
Sex of Examiner X Sex of Examinee		
Male Examiner-Male Examinee	100.44	6.46
Male Examiner-Female Examinee	103.12	4.92
Female Examiner-Male Examinee	101.50	5.73
Female Examiner-Female Examinee	102.50	5.63

TABLE 6 (cont'd.)

Effects	Means	Standard Deviations
Sex of Examiner X Expectancy		
Male Examiner-"Above Average" Examinees	102.94	5.32
Male Examiner-"Below Average" Examinees	100.62	6.22
Female Examiner-"Above Average" Examinees	102.81	6.02
Female Examiner-"Below Average" Examinees	101.19	5.23
Sex of Examinee X Expectancy		
Male Examinee-"Above Average"	102.44	5.75
Male Examinee-"Below Average"	99.50	6.07
Female Examinee-"Above Average"	103.31	4.88
Female Examinee-"Below Average"	102.31	5.64
Sex of Examiner X Sex of Examinee X Expectancy		
Male Examiner-Male Examinee-"Above Average"	102.00	6.48
Male Examiner-Male Examinee-"Below Average"	98.88	6.47
Male Examiner-Female Examinee-"Above Average"	103.88	4.09
Male Examiner-Female Examinee-"Below Average"	102.38	5.83
Female Examiner-Male Examinee-"Above Average"	102.88	5.46
Female Examiner-Male Examinee-"Below Average"	100.12	6.01
Female Examiner-Female Examinee-"Above Average"	102.75	5.80
Female Examiner-Female Examinee-"Below Average"	102.25	5.90

APPENDIX B

APPENDIX B

Rating Sheet

NAME _____

Group I.Q. Score
(OTMM-S Form)



School Achiev-
ment Average



Predicted WISC Score _____

APPENDIX C

APPENDIX C

TABLE 7

WISC TOTAL SCORES

Sex of Examiner	Examiner	Expectancy of Examiner by Examinee Sex			
		Male Examinee		Female Examinee	
		"Above Average" Expectancy	"Below Average" Expectancy	"Above Average" Expectancy	"Below Average" Expectancy
Male	1	107	106	95	101
		88	101	97	87
	2	116	101	101	90
		107	99	88	110
	3	111	88	83	101
		89	94	93	93
	4	100	89	109	89
		83	94	117	94
Female	5	98	106	107	91
		104	118	108	106
	6	103	105	93	104
		109	102	111	101
	7	112	109	85	107
		113	98	99	114
	8	110	91	99	88
		93	109	109	98

MICHIGAN STATE UNIV. LIBRARIES



31293104671148