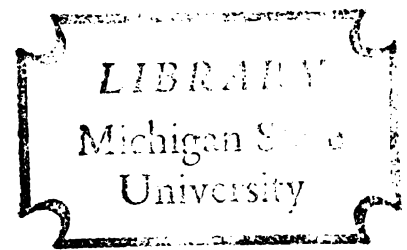


BIASED TEACHER REFERRALS AND THEIR EFFECTS
UPON SCHOOL PSYCHOLOGISTS' DIAGNOSTIC
JUDGMENTS AND EXPECTANCIES

Dissertation for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
DARRELL D. ELDERS
1977



This is to certify that the
thesis entitled
BIASED TEACHER REFERRALS AND THEIR EFFECTS
UPON SCHOOL PSYCHOLOGISTS' DIAGNOSTIC
JUDGMENTS AND EXPECTANCIES

presented by

DARRELL D. ELDERS

has been accepted towards fulfillment
of the requirements for

PH.D. degree in EDUCATIONAL PSYCHOLOGY
(SCHOOL PSYCHOLOGY)

Harvey F. Cloninger
Major professor

Date 2/9/77

~~P JUL 30 '79~~ 28

~~APR 5 '81~~ 063

~~APR 1981~~ 102

~~APR 1981~~ 102

ABSTRACT

BIASED TEACHER REFERRALS AND THEIR EFFECTS UPON SCHOOL PSYCHOLOGISTS' DIAGNOSTIC JUDGMENTS AND EXPECTANCIES

By

Darrell D. Elders

Two experiments were conducted to test the effects of biased teacher referrals upon diagnostic judgments and treatment recommendations of Michigan School Psychologists. School psychologists were exposed to one of three types of teacher referral reports, each describing the same child, but differing in the language used for this description. One referral described the child in negative terms such as "hyperactive, aggressive, and impulsive"; while another, positively oriented referral, described him in terms like "energetic, assertive, and spontaneous." A third behaviorally oriented referral reported only the child's observable behavior. The subjects studied identical information packets which contained the history and test protocols of an active six-year-old boy in first grade. All subjects viewed a video-taped recording of the boy in his classroom. Subjects rated the child's functioning on nine continua designed to

measure pathological severity. They also selected diagnostic labels from two lists taken from the Michigan "Mandatory Special Education Act" and clinical diagnostic classification systems, each of which included a psychologically healthy category. The subjects also selected a treatment from a list of special education interventions ranging from no special help at all to full-time placement in special education classrooms. Several independent variables were measured including the subjects' theoretical orientation (psychodynamic versus behavioral), length of professional experience, and type of graduate training (school psychology versus clinical psychology).

Four hypotheses were advanced, each predicting that the negatively biased teacher referral report would influence school psychologist subjects to over-estimate the child's pathology on one of four dependent measures: the "Severity of Pathology" scale, "Mandatory Special Education" diagnosis, clinical diagnosis, and the special education treatment the subjects decided to recommend.

The results of the main experiment, with 48 school psychologists from eight Michigan school districts, failed to show significant differences among the three types of teacher referrals on any of the four dependent measures. The same experiment, administered to a group of 12 school psychologists all employed by a medium-sized city school district, however, did result in significant ($p < .05$)

differences in the subjects' estimates of pathological severity. The negative referral condition produced significantly ($p < .02$) higher estimates of pathology than the positive referral condition, but not higher than the behaviorally oriented referral condition. This result could have been a function of the particular behaviorally oriented referral included in these experiments, or it could have been an indication that school psychologists experienced difficulty in evaluating behavioral data. This trend could also have been an indication that "halo effect" was slightly more powerful than negative expectancy effect.

Further analysis showed that the subjects' length of experience and type of graduate training were not highly related to their responses on the four dependent measures. These school psychologists' theoretical orientation showed moderate overall correlations with the dependent measures, but the values were lowest, not highest, in the negative teacher referral condition. Although these experiments were not specifically designed to test for school district effects, the judgments of school psychologists were significantly different among the districts sampled ($p < .007$).

It was concluded that Michigan school psychologists are unlikely to be inappropriately influenced by

biased teacher referrals, at least in so far as their judgments would substantially affect the life and school experience of children. The effects of the school psychologists' school district upon their diagnostic judgments and treatment recommendations appeared to be worthy of further study. Discovery of which variables, within a school district, might account for these effects was deemed to be particularly important. Further study of the effects of behaviorally oriented pre-evaluation information was also recommended.

BIASED TEACHER REFERRALS AND THEIR EFFECTS UPON
SCHOOL PSYCHOLOGISTS' DIAGNOSTIC JUDGMENTS
AND EXPECTANCIES

By

Darrell D. Elders

A DISSERTATION

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

1977

DEDICATION

To Pat, my wife and friend, and to my children,
Randy, Barbie, Jeannie, and Becky, who have so lovingly
sustained and encouraged me.

ACKNOWLEDGMENTS

The members of my committee have not only given me the guidance and friendship necessary for my professional development, but their influence has and will be felt in my personal development as well. I am honored to acknowledge them here with my gratitude.

--To Harvey Clarizio, my chairman and teacher, whose training has changed my life, and whose creative thought, knowledge of the field, and personal example are evident throughout this study.

--To Robert Craig, who, despite his heavy responsibilities and demanding schedule as department chairman, unselfishly gave many hours of study and guidance in the final preparation of this study.

--To J. Edwin Keller, for his many helpful suggestions in the planning stages of this study and for his encouragement and warm friendship during its completion.

--To William Hinds, whose ideas and knowledge were especially helpful in the early stages of this study.

TABLE OF CONTENTS

Chapter	Page
I. THE PROBLEM	1
Need for the Study	5
Purpose of the Study	8
Overview of the Study	11
II. REVIEW OF THE LITERATURE.	12
Teacher Expectation Studies	13
Experimenter Expectancy Effects.	24
Sex	26
Prestige.	27
Experimenter Experience.	27
Anxiety and Need for Approval.	28
Experimenter-Subject Acquaintance	29
Task Variables.	30
Mediating Variables	31
Psychometric and Clinical Experiments.	33
Intelligence Test Expectancy Studies with Examiner-Examinee Interaction	35
Intelligence Test Expectancy Studies without Examiner-Examinee Inter- action	40
Expectancy Effects in Personality Testing	47
Interview, Observational and Assessment of Therapy Experiments	52
Labeling Bias in Clinical Studies	55
Summary and Implications for this Study	61
III. METHODOLOGY	66
The Sample.	66
Experimental Stimuli	68

Chapter	Page
Teacher Referral Report	70
Background Information Sheet	71
Revised Wechsler Intelligence Scale for Children Protocol	72
Bender Visual Motor Gestalt Test Protocol	72
Kinetic Family Drawing	73
Wide Range Achievement Test Protocol. .	73
Video-Taped Behavior Observation . . .	74
Dependent and Independent Measures . . .	75
Severity of Pathology Scale.	75
Diagnostic Classification	77
Degree of Confidence Scale	78
Special Education Treatment Recommen- dation	78
Referral for Pediatric Determination of Medication Possibilities	79
Theoretical Orientation	79
Experience and Training	80
Procedures	80
Sampling Procedure.	80
Experimental Procedure	81
Experts' Judgments.	84
Hypotheses	86
Design and Analyses	89
IV. ANALYSIS AND DISCUSSION OF RESULTS. . . .	93
Analyses for Hypotheses.	93
Analyses of Relationships and Associ- ations.	95
Length of Experience	99
Theoretical Orientation	101
Confidence	102
"Severity of Pathology"	103
Special Education Treatment Recom- mendation	104
"Classical" Diagnosis, "Mandatory" Diagnosis and Referral for Medication.	105
Graduate Training in School Versus Clinical Psychology.	107

Chapter	Page
School District	109
Grand Rapids Study.	111
Summary and Discussion of Results. . .	117
V. SUMMARY, CONCLUSIONS, AND IMPLICATIONS. .	132
Summary	132
Conclusions	137
Implications for Practice and Future Research	142
APPENDICES	
APPENDIX	
A. TEACHER REFERRAL REPORTS	149
B. BACKGROUND INFORMATION SHEET	152
C. WECHSLER INTELLIGENCE SCALE FOR CHILDREN-REVISED PROTOCOL	153
D. BENDER VISUAL MOTOR GESTALT TEST PRO- TOCOL	159
E. KINETIC FAMILY DRAWING	161
F. WIDE RANGE ACHIEVEMENT TEST PROTOCOL . .	162
G. SEVERITY OF PATHOLOGY SCALE	163
H. RATING SHEET: DIAGNOSES, CONFIDENCE, SPECIAL EDUCATION TREATMENT RECOMMENDATION	164
I. RATING SHEET: REFERRAL FOR MEDICATION, THEORETICAL ORIENTATION, TRAINING EXPERIENCE	165
J. RAW DATA FREQUENCY DISTRIBUTIONS. . . .	166
K. RANDOM GROUP TESTS OF ASSOCIATION . . .	168
REFERENCES	172

LIST OF TABLES

Table	Page
3.1 Sample Characteristics	69
4.1 Random Group Univariate Analysis of Variance	94
4.2 Random Group Cell Means and Standard Deviations	95
4.3 Random Group Correlations, Associations, and Significance Levels with all Treat- ment Groups Combined	96
4.4 Random Group Correlations, Associations, and Significance Levels within the Nega- tive Treatment Condition.	97
4.5 Random Group Correlations, Associations, and Significance Levels within the Positive Treatment Condition	98
4.6 Random Group Correlations, Associations, and Significance Levels within the Behaviorally Oriented Teacher Referral Condition	100
4.7 Analyses of Variance for Graduate Training .	108
4.8 School District Groups Analyses	110
4.9 Grand Rapids Group Univariate Analyses of Variance	112
4.10 Grand Rapids Group Cell Means and Standard Deviations	113
4.11 Grand Rapids Group Correlations, Associa- tions, and Significance Levels with all Treatment Groups Combined	114
4.12 Grand Rapids Group Fisher's Exact Tests with all Treatment Groups Combined	115

Table		Page
J-1	Raw Data Frequency Distributions	166
K-1	Random Group Chi Square Tests of Association with all Treatment Conditions Combined	168
K-2	Random Group Fisher's Exact Tests Within the Negative Treatment Condition	169
K-3	Random Group Fisher's Exact Tests Within the Positive Treatment Condition	170
K-4	Random Group Fisher's Exact Tests Within the Behaviorally Oriented Teacher Referral Condition	171

LIST OF FIGURES

Figure	Page
3.1 Experimental Design	90

CHAPTER I

THE PROBLEM

Since its inception as a basic diagnostic service for the purpose of determining levels of mental retardation in school children, school psychology has expanded and refined its expertise to a level of considerable sophistication. Special education remedial programs have concurrently broadened and specialized so that both diagnosis and treatment are available to more children than ever before. As of 1970, The Bureau of Education for the Handicapped reported over 1,700,000 children being served in special education programs for the retarded, emotionally handicapped, visual and hearing impaired, and learning disabled. Continued expansion is assured by the estimated prevalence of these disorders in over four million school age children (Hobbs, 1975, pp. 50-79).

Along with increased numbers of children coming under psychological scrutiny has come growing professional and lay concern with the issue of labeling and classification. It is feared that labels can have a stigmatizing effect upon their recipients. It is commonly believed

that labels tend to produce expectancy sets on the part of those dealing with the child so that areas of opportunity can become limited or closed. Perhaps the most well-known experiment purporting to demonstrate expectancy effect was reported by Rosenthal and Jacobson (1966) in which teachers were led to expect that randomly labeled "intellectually blooming" students would increase in their ability. The authors' results showed that these same students made more gains on post IQ tests than students not so labeled. Many studies have demonstrated that labels such as mentally retarded are associated with low expectations and negative stereotypes and attitudes (Boekel, 1969; Guskin, 1963a and b; Hollinger & Jones, 1970; Jones, 1972; Meyers, Sitkei, & Watts, 1966; Shotel, Iano, & McGettigan, 1972). These stereotypes can be so powerful for teachers that when children exceed the expectations set for them, they are often seen as having problems with social adjustment (Rosenthal & Jacobson, 1968). Expectancy effects can be quite resistant to educational measures taken to extinguish them. Foster, Ysseldyke, and Reese (1975) found that even when their sample of graduate and undergraduate students in special education were exposed to the work of Rosenthal and Jacobson, as well as other expectancy and bias studies, they continued to hold negative, stereotyped expectations for the children they observed.

The Rosenthal-Jacobson findings have been called into serious question on methodological grounds (Elashoff & Snow, 1971; Snow, 1969; Thorndike, 1968) and for failures to produce like results in subsequent replications (Clai-born, 1969; Fleming & Anttonen, 1971a and b; Jose & Cody, 1971). As will be shown later, however, there remains enough evidence in the area of negative teacher expectations to create apprehension among most educators, including school psychologists. In fact, many school psychologists, fearing expectation effects, attach statements such as "he shows higher intellectual potential," or "this should be considered a minimal estimate of her ability" to their evaluation reports of lower ability children.

Although the practice of labeling is a source of concern, experience confirms that it continues to exist. In fact, diagnostic labeling has become more prevalent than ever before. Part of the reason for the ubiquity of labels most certainly accrues from the fact that they are an efficient means of transmitting and conceptualizing large amounts of information. Indeed, much of the information accumulated regarding handicapped children has been organized around such categories. Even more important, however, is the fact that needed help often cannot be provided children without the application of a label of some kind. The federal government has appropriated funds for the education of handicapped children (Education

for all Handicapped Children Act, Public Law 94-142, 1975), but requires that such children be formally labeled. This law defined handicapped children as

. . . mentally retarded, hard of hearing, deaf, speech impaired, visually impaired, or otherwise health impaired children, or children with specific learning disabilities who by reason thereof require special education and related services. (Section 602, paragraph 1)

In Michigan, the Mandatory Special Education Act (Public Act 198, 1971) not only provides funds for the education of handicapped children, but makes it mandatory that all handicapped children in the state receive a diagnosis and special education services. Since federal law (P.L. 94-142) requires the states to implement a state-wide plan for the delivery of special education services, it is likely that all states will enact legislation similar to Michigan's within the next few years. Indeed, several states have already done so. As a consequence, the practice of diagnostic labeling is almost certain to continue for some time, at least as it is practiced in the schools.

Given the fact that one will apparently live with the possible negative consequences of labeling in order to obtain needed help for exceptional children, the occurrence of misdiagnosis or inappropriate labeling is especially abhorrent. To suffer the stigma attached to an accurate diagnosis is one thing, but to endure the stigma attached to misdiagnosis is quite another. Increased quality and quantity of professional training

have been the preferred approaches to a cure for the problem of misdiagnosis. Unfortunately, even the best diagnostician is only as good as the information he receives. Furthermore, it is possible that the diagnostician is just as susceptible to expectancy set produced by pathological labels as are the consumers of the diagnoses he makes.

Need for the Study

Diagnostic error resulting from biased referral information can be a problem in psychology and psychiatry. Although several hundred studies have been reported in the general area of expectancy set or bias, the literature is quite silent with regard to clinical diagnostic error resulting from bias in psychological or psychiatric evaluations. If one were to delete those studies which have used graduate school students as subjects and those studies which have focused upon intelligence or personality test scoring as the dependent variable, published data in this area of interest would be very sparse indeed.

In school psychology, published expectancy studies are confined completely to test scoring and examiner-examinee test behavior experiments. Despite the fact that school children are now probably the largest single group of recipients of psychological services in the United States, they receive these services from a group of professionals who have been given little attention

in the experimental literature and about whom little is known. The judgments of school psychologists can have a profound effect upon school children throughout their educational careers and the evidence is that ever-increasing numbers of children are exposed to these judgments. The possible effects of bias upon these judgments is therefore of critical concern.

In the mental health fields diagnosis has generally focused on the client's negative characteristics, or deviant behavior, rather than emphasizing the individual's positive characteristics. The clinician practicing with the illness orientation, then, tends to diagnose pathology when he is in doubt, rather than risk overlooking a possible disease condition which may be harmful to his patient. Despite various attacks made upon the medical model (Sarbin, 1967; Szasz, 1961), school psychologists, no less than other mental health professionals, have been trained with materials espousing this same orientation and thus continue to focus most of their attention upon pathology. Underlying most diagnoses, then, there appears to be a general bias toward pathology.

In addition to this general pathological bias, school psychologists are exposed to certain unique pressures which may serve to increase the strength of this bias. With the proliferation of special education programs and the need to have full classrooms in order to

qualify for reimbursing funds, school psychologists are under frequent pressure from their administrations to fill these special classrooms quickly. Classroom teachers have also been known to exert pressure upon school psychologists to diagnose handicapping conditions so that troublesome children can be removed from their classrooms. Mandatory Special Education legislation in Michigan and some other states has also alerted many parents to the variety of special help possibilities for their children and these laws have allowed parents the vehicle to exert considerable pressure upon special education administrators and school psychologists to obtain this help, even for mild learning problems. The area of learning disabilities has been especially problematic in this regard. Parents of children with learning disabilities have formed powerful and vocal groups with influence at all levels of school administration. They are well informed with regard to schools' legal responsibilities and are not hesitant to press for special services. This pressure, coupled with the nebulous quality of the handicap, has often produced situations in which it is easier for the school psychologist to diagnose learning disability than to avoid the label.

The referral agent for most school psychological evaluations is the classroom teacher. It is the teacher who makes the first judgments concerning a child's behavior, and who often transmits these to the school

psychologist within the referral. The psychological evaluation is not an isolated activity, as Towbin (1960) points out; it consists of a triadic group of individuals, the patient, the referring person, and the examiner, who mutually influence each other. The referral is made because the referring person expects that the psychological evaluation will serve his own purpose in some way. Within the bias and expectancy literature, appallingly little has been published concerning the influence the classroom teacher has upon the school psychologist's diagnostic judgments.

Purpose of the Study

The purpose of this study was to add knowledge to existing and somewhat meager information concerning the effects of biased referrals upon the psychological evaluation of children. Specifically, this experiment focused upon judgments made by school psychologists relative to the diagnosis of learning disabilities in children. The study was designed to enhance knowledge in this area by going a step further than previous studies which have measured differences in test scoring as a dependent variable to bias induction. It could be argued that even moderately large deviations in test scores resulting from bias may not be enough to influence a child's final diagnosis and what will subsequently happen to him. Consequently, this study attended primarily to school

psychologists' judgments concerning diagnostic classification, severity of the pathology, and special education treatment as dependent variables with the test scoring element held constant.

To add to the generalizability of the findings, this study used actual practicing school psychologists as subjects, rather than students in testing courses as has been done in many previous experiments. The reactions of these subjects were studied in relation to their theoretical orientation and length of professional experience.

The effects of biased teacher referrals in this experiment were studied in the context of evaluating a possible learning disabled youngster. Attention was paid to all aspects of his functioning, not just his intelligence. It is hoped that the results of this study will, therefore, have broad applicability and that its variables will have appropriateness to the current practice of special education referral and school psychology.

This study was designed to test the hypothesis that positively and negatively biased teacher referrals differentially affect judgments made by school psychologists in their evaluation of a possible learning disabled boy. It was expected that these effects could be observed in school psychologists' willingness to label the child with a pathological classification, their estimate of the severity of this pathology, and, as a consequence,

their judgment concerning the amount of special education intervention necessary to remediate or manage the problem. Formal statements of four specific hypotheses are presented in Chapter III.

Further, since any school psychologist brings to his work an orientation toward diagnosis that is dependent in part upon his training and experience, it might be expected that these variables are related to his product: judgments concerning adjustment, diagnosis, and necessary treatment. Some evidence has suggested that theoretical orientation and length of experience may be related to biasability. Behaviorally oriented psychologists may be less susceptible to bias than their psychodynamically oriented colleagues (Langer & Abelson, 1974). Less experienced clinicians may also be influenced somewhat less by pre-diagnostic information than experienced workers in clinical situations (Temerlin, 1968). It was anticipated, therefore, that the experience and theoretical orientation of this study's subjects would show significant positive correlations and associations with their diagnosis, their estimates of pathological severity, and their recommendations of special education intervention when they were exposed to negatively biased information concerning the subject of the evaluation. Although this study was not designed to provide an experimental test of these expectations, data were collected from the school

psychologists sampled which allowed for an exploration of relationships between the above variables. These relationships, while being descriptive of the sample only, were analyzed for the purpose of determining whether the variables would merit further study.

Overview of the Study

Chapter II, next, will begin with a discussion of the types of studies previously done in the general area of expectancy. Several hundred experiments have been carried out in this general area, many of which are only tangentially related to the variables of interest in this experiment. For this reason, only those studies which have a rather direct bearing upon the process of diagnosis as it relates to expectancy and bias will be reported and analyzed in depth.

The experimental design, including sample description and sampling method, instruments, procedure, statistical hypotheses, and plan for statistical analysis will be presented in Chapter III. The statistical analysis and discussion of the results of the experiment will be the focus of Chapter IV. Lastly, Chapter V will contain a summary of the study and will present the reader with a discussion of conclusions and implications for practice. Suggestions for future research will also be offered.

CHAPTER II

REVIEW OF THE LITERATURE

The experimental literature pertaining to bias and expectancy is truly voluminous. In his review of the literature on expectancy effects, Rosenthal (1976) observed that there were well over 300 experiments reported in the field. Most of these studies were concerned with the effects of expectations in general experimental situations and were focused primarily upon differential subject behavior as a result of experimentally induced expectations given to their experimenters. Manipulations in experimental situation, task, experimenter, and subject variables comprised the subject of most of these studies.

Besides the above array of literature, special applications of bias and expectancy experiments have built up a formidable volume of research in their own right. Of particular relevance to this study are those experiments reported in the field of education with teacher expectations, and in the field of psychology with examiner bias studies in testing situations. The

review of related literature in this chapter begins with a summary of the major findings in the area of teacher expectancy, together with citations of representative experiments, and following, a discussion of some pertinent knowledges gleaned from the broader area of experimenter expectancy research. The chapter will conclude with a more detailed review of reported experiments in the areas of psychometric and clinical evaluation.

Teacher Expectation Studies

In 1966, Robert Rosenthal and Lenore Jacobson published the results of an experiment in which 20% of the children in each of 18 classrooms were randomly reported to their teachers as showing unusual potential for intellectual gains. This potential was said to be shown on a test designed to predict academic "blooming," but was really the Flanagan Tests of General Ability (1960), a nonverbal intelligence test. Eight months later, these "blooming" students showed significantly greater gains on the same IQ test, administered a second time, than did the remaining children in the control group. Although the study included children in grades one through six, these expectancy effects were significant only for the first two grades.

Two years later, Rosenthal and Jacobson (1968) published a more complete discussion of the 1966 "Oak

School" experiment together with a presentation of supporting research and implications. This book, Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development, swiftly became one of the most controversial works in educational literature and touched off a flurry of subsequent teacher expectation experiments.

The Rosenthal and Jacobson (1968) findings have been subjected to severe criticism on methodological grounds. Thorndike (1968) attacked the "Pygmalion" study on the basis of reported pre- and post-test scores. He showed that many of these scores could not possibly have occurred, and therefore the entire study is based upon an analysis of faulty and valueless data. In his short but scathing review, Thorndike (1968, p. 708) says of the study, "Alas, it is so defective technically that one can only regret that it ever got beyond the eyes of the original investigators!" Jensen (1969) showed that Rosenthal and Jacobson's use of the same IQ test for pre- and post-measures, use of teacher-administered tests, and use of the child as the unit of analysis rather than the classroom, all work together to cast considerable doubt upon their results. Claiborn (1969) criticized the Rosenthal-Jacobson data analysis for failure to correct for known pre-test differences, for practicing partial data analysis, and for probability pyramiding through multiple significance tests.

Although Rosenthal and Rubin (1971) have answered many of the above criticisms, the debate concerning the Rosenthal and Jacobson research continues, and, in fact, is the subject of a full-length book (Elashoff & Snow, 1971). Despite the criticisms left unanswered and the questionable validity of Pygmalion in the Classroom, its central hypothesis, that teachers' expectations for students can become a self-fulfilling prophecy, appears to have real merit. In addition, many reviewers have pointed out that the work of Rosenthal and Jacobson should be valued for its effect in prompting other, more sound, research in this area.

The number of studies published in the area of teacher expectancy is now in excess of 60. Even the firmest believer in teacher expectancy effects will be forced to admit that none of these studies has unequivocally replicated the original Rosenthal and Jacobson study and findings. To be sure, expectancy effects have been found, but not in exactly the same way or with exactly the same measures as were used in the "Oak School" experiment. An exhaustive review of the more than 60 published teacher expectation studies will not be presented here. A more thorough review can be found in Baker and Crist (1971) and, more recently, in Brophy and Good (1974) and Dusek (1975).

One of the greatest obstacles to demonstrating teacher expectancy effects is the fact that the induction of expectancy is not always successful in an experiment. Brophy and Good (1974) have analyzed a number of failures to replicate Rosenthal and Jacobson's findings and conclude that many of the experimenters were simply not successful in getting their teachers to believe or remember what they were supposed to believe about their students. In support of this conclusion, Brophy and Good cite several examples where large percentages of the subjects failed to remember which of their students were "bloomers," or disregarded the expectancy induction because they guessed the purpose of the experiment. Further, if one tallies the number of positive findings in experimentally induced teacher expectation studies versus those in naturalistic studies, where teachers have already formed their own expectancies, the majority of positive findings in the naturalistic experiments shows the importance of successful induction.

The variables of professional experience and teacher-student acquaintance appear to be important to the success of teacher expectation experiments. Most studies which have demonstrated expectancy effect in teachers have used graduate or undergraduate students with little or no experience in teaching. Rubavits and Maehr's (1971) study using undergraduate volunteer

teachers and Carter's (1969) experiments with University of Indiana students are examples of studies which have shown powerful teacher expectation effects with inexperienced teachers. The Rothbart, Dalfen and Barrett (1971) study with teacher trainees also showed differential treatment of students depending upon teacher expectations.

In some studies the expectation that certain students were of "high" ability simply did not match the teacher's perceptions of that student. This is perhaps one of the reasons why studies begun late in the school year have failed to produce the experimentally desired expectations in teachers. In these situations the teachers have had enough time to form their own impressions of students before the beginning of the experiment. The negative findings of the Jose and Cody (1971); Conn, Edwards, Rosenthal, and Crowne (1968); Fielder, Cohen, and Feeney (1971); and Claiborn (1969) studies are good examples of experimentally induced teacher expectations which did not occur until the second semester or later. On the positive findings side of the ledger, studies in which the student and teacher did not know each other seem to find the process of induction easier. The Carter (1969); Rothbart, Dalfen, and Barrett (1971); and Rubavits and Maehr (1971) studies, cited earlier, all showed expectation effects with teachers who were not acquainted with their students. Brown's (1969) doctoral dissertation

also showed expectation effects with subjects and students who were not acquainted. Rosenthal and Jacobson's (1966, 1968) original study showed significant IQ gains for the first two grades only ($p = .002$ for the first grade and $p = .02$ for the second) and with more gains for the first graders than the second. Presumably the first grade teachers were acquainted with fewer of their students than teachers in the later grades. The most spectacular expectancy effects are also observed in tutorial situations where students and their tutors have not been acquainted before the experiment. Unfortunately, most studies of tutorial teaching situations also involve inexperienced teachers. Thus, with the variables of experience and acquaintance confounded, it is difficult to separate their effects and significance.

In their review of the teacher expectancy literature, Brophy and Good (1974) distinguish between process and product measures used as criterion or dependent variables. Product variables are those which show the outcome of the self-fulfilling prophecy such as differences in IQ or achievement test scores. Process variables are those intervening teacher behaviors which are suspected of causing differences in learning. When one examines the literature with the distinction of these two types of variables in mind, abundant evidence of differential treatment, or process results appear as a

consequence to different teacher expectancies. The experiments of Medinnus and Unruh (1971); Rothbart, Dalfen, and Barrett (1971); and Rubavits and Maehr (1971) all suggest that teachers tend to interact more frequently and more positively with high expectation than with low expectation students. Brophy and Good (1970) found that teachers communicated their expectations of students by demanding more of high expectation students and rewarding them more frequently. Low expectation students, on the other hand, were the butt of more frequent criticism, received less feedback for answers, and were given fewer chances to respond. In another example of process variable expectation effect, Chaikin, Sigler, and Derlega's (1972) sample of undergraduate tutors smiled three times more often, nodded their heads two and one-half times more often, looked their students in the eye more often, and leaned toward them longer when they thought they were working with bright students than when their expectation was of either normal or dull intelligence.

The successful induction of differential expectations and even the differential treatment of students as a result of this induction does not necessarily lead to differential product measures such as the tested learning of a child or his IQ. Brown's (1969) doctoral dissertation is a good example of this phenomenon. In

this tutoring situation involving the learning of states and their capitals, Brown found that his high expectation students learned no more than controls, but the teachers with high expectations did attempt to teach more than did teachers with low expectations. Kester and Letchworth (1972), in their study using seventh grade students, also found no expectancy advantage in product measures (achievement, IQ, and aptitude tests), but process observations showed that teachers spent more time with high expectancy students, talked with them more often, and were generally more supportive toward them.

The length of time that teachers and students are together in an experiment seems to have some bearing on the outcome. Generally, those studies which have spanned less than a full academic year have been more successful in showing teacher expectancy effect than those studies lasting for longer periods of time. It is suspected that this is true because longer periods of exposure to their students allows teachers a chance to form their own independent notions regarding the ability of these students and allows for more time to forget the experimenter's induced expectancies. Perhaps the Conn, Edwards, Rosenthal, and Crowne (1968) experiment provides the most dramatic example of this observation. The Conn et al. study was a partial replication of the original Rosenthal and Jacobson (1966, 1968) "Oak School"

experiment using the same basic paradigm. Two sets of data were taken, one at four months and another at three semesters after expectancy induction. While the first set of data showed small differences in favor of the high expectancy students, the second set actually reversed this state of affairs so that in the end, the control group showed the advantage to a degree very near statistical significance. The self-fulfilling prophecy hypothesis is most powerfully supported by the experiments of Meichenbaum, Bowers, and Ross (1969) and Beez (1968). Both of these experiments showed significant expectancy effect on both product and process measures. The Meichenbaum et al. experiment lasted for two weeks and Beez's subjects tutored students for 15 minutes.

As Brophy and Good (1974) suggest, teacher expectancies are not necessarily harmful. Most expectancies formed are accurate and help the teacher adjust his instruction to the individual needs of the child. Beez's (1968) study, for example, showed that "low ability" children were given more examples and explanations and were taught at a slower pace than "high ability" students. If those children were actually low ability students, their instruction would have been entirely appropriate for them. Even if the teacher's initial expectations are in error, most professional teachers quickly adjust these expectations after having worked with the child. It is

only when teachers rigidly hold to their initially false expectations for students that potentially harmful conditions ensue. This kind of rigidity is usually the result of poor teaching or poor mental health. Despite the fact that inappropriate expectancy set is an example of bad teaching, it does continue to exist as is so convincingly demonstrated by Rist (1970) who cites anecdotal evidence of the inappropriate "caste system" placement of children whose expectations were formed arbitrarily by kindergarten teachers and maintained by later primary grade teachers.

In summary, the findings of the original Rosenthal and Jacobson (1966, 1968) study have not been replicated by any full academic year experiment using IQ or achievement as the dependent variable in an experimentally induced expectancy design. In fact, the experimental literature has not generally produced convincing evidence that teachers' expectations have a differential effect upon the actual learning or intellectual development of children. Examples of the difficulties involved in expectancy induction were presented which showed that the expectancy must be believable and must not be at variance with impressions the teacher has already formed of the child in question. Therefore, experiments which focus on students and teachers who are not acquainted with each other generally show more positive results.

This is also perhaps one of the reasons why positive results have been found with younger children and in tutorial situations where teachers and students meet for the first time. Experiments of short duration also tend to show more positive results, again perhaps because the induced expectation remains accepted by the teacher who has less time to interact with the child and to form conflicting opinions regarding his abilities. Studies which use novice or relatively inexperienced teachers also report positive expectation results more frequently than those using experienced professionals. Presumably the inexperienced teachers are more open to suggestion and are less experienced in detecting student behavior that is contradictory to the induced expectation.

That teachers do form expectations for students and that these expectations can produce differential teaching behavior is convincingly demonstrated by the majority of studies which focus on process variables in addition to, or instead of, the product measures of IQ and achievement test scores. Often, however, children are resistant to the differential treatment of their teachers; several studies have shown teacher expectation effects on process variables such as warmth, frequency of interaction, and reinforcement patterns, but have failed to show the effects of these behaviors on objective tests of achievement or IQ.

Although the literature on experimentally induced teacher expectation effects is sometimes contradictory, ambiguous, and confusing, studies which have been concerned with the effects of teachers' own, naturalistically formed expectations are very often cited as the most convincing evidence of the self-fulfilling prophecy. Most naturalistic studies are contained in the literature on tracking which has shown, quite universally, that teachers' expectations result in the differential treatment of children. High expectancy children are usually treated appropriately while low expectancy children are not, and are, therefore, adversely affected. Moreover, once a child is placed within a track or expected ability group, his chances of remaining in that group, regardless of his true ability, are surprisingly high. Unfortunately, it is difficult to understand clearly the results of tracking since outcome or product measures in these naturalistic studies may have been true measures of the students' abilities and their teachers may have responded to real student differences rather than false or biased expectations.

Experimenter Expectancy Effects

The classic paradigm for studying experimenter expectancy effects is the "person perception task" designed and standardized by Robert Rosenthal and first reported by Rosenthal and Fode (1961). This task

involves the presentation of 10 previously judged neutral photographs of faces to subjects who are asked to rate the people shown in these photographs as experiencing success or failure on a scale from -10 to +10. In the initial experiment using the person perception task (Rosenthal & Fode, 1961), five experimenters were led to expect an average rating of +5 from their subjects while another five were led to expect average ratings of -5. The results indicated that experimenters anticipating an average rating of +5 in their subjects actually obtained a mean rating of +.4 while the five experimenters who expected an average of -5 actually obtained a mean rating of -.08 ($p < .007$). As part of his master's thesis, Kermit Fode replicated this experiment and published his results with Rosenthal in 1961 and again in 1963 (Rosenthal & Fode, 1961, 1963). His results were more impressive and showed mean ratings of +2.27 and +.48 for the +5 and -5 expectancy treatment conditions ($p < .0003$).

Since the original Rosenthal and Fode experiments, over 100 studies have been published using the person perception task as the vehicle for examination of various aspects of experimenter expectancy. In his most recent examination of the literature, Rosenthal (1976, p. 442) found that person perception studies yielded significant ($p < .05$) results 25 to 30% of the time. By simple head

count, therefore, one would be forced to conclude that the balance of evidence weighs most heavily toward a judgment that experimenter effects in this area are not significant and are difficult to demonstrate. Further, Rosenthal's (1976, p. 442) analysis of over 300 studies including experiments in reaction time, ink blot tests, animal learning, laboratory interviews, psychophysical judgments, learning and ability, person perception, and everyday situations yielded significant ($p < .05$) expectancy effects in only 35% of the cases. However, Rosenthal (1976, pp. 441-442) argues that significant findings in the approximately one-third of published expectancy studies is a significant finding overall since this proportion is about seven times what would be expected by chance alone.

Most experimenter expectancy studies have not really concentrated upon the question of whether or not these effects occur, but rather have been concerned with questions such as under what circumstances and with whom are these effects most powerful, and how are the effects mediated? Several of these variables are discussed below.

Sex

The literature does not clearly settle the question of sex influences in experimenter expectancy. First, the studies are conflicting, some showing greater male influence, some female, and some showing no difference.

Further, there seems to be a difference in experimenter effect depending upon the sex of their subjects and the type of task in which they are involved. Finally, the experimenter's sex variable has often been confounded with other variables such as the experimenter's prestige or dominance. Rosenthal (1969a, pp. 240-242; 1976, pp. 225-233), however, believes that there is some overall evidence to suggest that male experimenters tend to produce the strongest expectancy effects. These effects become more pronounced when male experimenters interact with female subjects.

Prestige

The prestige or status of the experimenter has been a variable of interest in several studies. In general, experimenters of higher status tend to be better able to influence their subjects in expected directions. However, the status of the experimenter is often confounded with other variables such as age, "air of authority," and simple interpersonal competence in most studies. The prestige variable may not be important in itself, but may depend more upon some of the reasons why certain people become prestigious.

Experimenter Experience

As was the case in the teacher expectation literature summarized earlier, the greatest effects in general



experimenter expectancy studies are usually found with less experienced experimenters. The experience variable may also be indirectly examined in experiments which cover a long period of time. It could be assumed that in such experiments, the experimenter gains experience as time progresses. Sampling of experimenter effects at progressive intervals has shown that expectancy effects are more powerful at the beginning of an experiment than during its later stages.

Anxiety and Need for Approval

The anxiety of the experimenter has been found to be important in several expectancy studies. Anxiety levels in both experimenter and subject have been shown to be significantly related to expectancy effects, but the direction of this relationship has been remarkably unpredictable. It appears that highly anxious experimenters may show their greatest expectancy effects with subjects who are also highly anxious. Conversely, when low-anxious experimenters show their greatest expectancy effects, their most susceptible subjects are likely to have less than average anxiety as well.

Several studies have shown that expectancy effects were positively related to the experimenter's need for social approval, but only in the case of experimenters whose level of anxiety was about average. When the experimenter had either high or low anxiety, the

relationship between expectancy effects and experimenter need for approval appears to have reversed itself and minus correlations were observed. Although it would be expected that subjects with a high need for approval would show greatest susceptibility to expectancy influence, studies have consistently shown that no relationship exists between these two variables. It appears, then, that need for approval has more to do with a person's ability to influence someone else than to their susceptibility to this influence. This influence occurs in expected directions when the experimenter's anxiety level is about average, but when anxiety is either high or low, the experimenter's influence produces opposite than expected results.

Experimenter-Subject Acquaintance

Teacher expectation studies, it may be remembered, have tended to show greatest expectation effects in situations where the teacher and student were not acquainted. The suspected reason for this phenomenon is that as the teacher gained a better knowledge of the student, the induction of the intended expectancy tended to weaken. Teachers began to form their own expectancies based on their experience with the child rather than depending upon the artificially induced expectation. In the experimenter expectancy literature this situation appears to be reversed. Here, greatest expectancy



effects have usually been obtained when subjects have known their experimenters. Examination of the experimenter expectancy studies, however, shows this finding only in the case of male experimenters. When female experimenters were used, the results were similar to the findings in the teacher expectancy literature. The discrepancy between teacher and experimenter expectancy studies on this variable, then, may be more a reflection of sex differences than anything else. It is probable that most teacher expectancy studies used female instructors while most experimenter expectation studies have used male experimenters.

Task Variables

In answer to criticism by Dana and Dana (1969) that the evidence for experimenter expectancy effects comes almost entirely from person perception experiments, Rosenthal (1969b) reviewed some 40 experiments which employed other tasks. In general, Rosenthal noted, tasks other than person perception tend to show the greatest effects of experimenter expectancy. Sixty-six percent of his 40 study sample showed significant ($p < .10$) experimenter expectancy effects as compared with only 36% significant ($p < .10$) expectancy effects obtained in person perception task studies.

In a more complete and recent breakdown of task variables related to expectancy effects, Rosenthal (1976,

p. 443) verified his 1969 findings, this time at the more acceptable 95% confidence level. Expectancy experiments in animal learning showed a 64% significant ($p < .05$) effects finding, while person perception tasks showed only 27% significant effects above the .05 chance level. In fact, of the eight types of experimental tasks reviewed and compared, person perception task experiments were tied with learning and ability expectancy studies for the lowest position of the group.

Several authors have argued that relatively structured or factual tasks may show less experimenter expectancy effect than ambiguous or less structured tasks. This postulation has been shown directly in a study by Shames and Adair (1967) and is supported by several other experiments.

Mediating Variables

Barber and Silver (1968) have outlined 11 possible ways experimenters' expectancies can influence the results of their research. Experimenters may unintentionally influence their subjects to give expected responses through paralinguistic cues (e.g. variations in voice tone), kinesthetic cues (e.g. changes in facial expression), or verbal reinforcement of desired responses. They may also intentionally or unintentionally misjudge or misrecord subject responses. Finally, they can simply fabricate their data. In Rosenthal's (1966, 1976) reviews

of the literature, the focus was primarily upon unintentional influence through paralinguistic and kinesthetic cues. In a careful review of 12 studies showing apparently significant effects of expectancy, Barber and Silver (1968) found that these effects were most commonly produced by errors in judging, recording, or reporting the experimental results. Verbal reinforcement as well as paralinguistic and kinesthetic cues also communicated the experimenters' desires in several of these studies. Barber and Silver were led to suspect that when student experimenters test student subjects on relatively structured tasks, the most prominent mode of expectancy mediation may be intentional data misreporting by the experimenters. In situations where experimenters test subjects on relatively ambiguous tasks, unintentional transmission of expectancy through paralinguistic and kinesthetic cues may be the most prominent form of mediation.

In summary, it must be admitted that perhaps the experimenter expectancy literature is more impressive for its bulk than its incidence of significant findings. A "box score" of significant expectancy effects ranges from a high of 64% of studies in the area of animal learning to a low of 27% of those experiments conducted in human learning and ability and with person perception tasks. The phenomenon of expectancy effect is not so simple as to show itself without complication. Expectancy effects

are modified by and show interaction with such variables as sex, experimenter and subject personality characteristics, experimenter status and experience, and the degree to which the experimenter and subject are acquainted. The type of experimental task appears to make a difference in expectancy effects and the responses of early subjects seem to exert an influence upon the expectations of the experimenter, as well. Although Rosenthal (1976) believes that experimenter expectancies are transmitted to subjects through such subtle cues as the experimenter's voice tone or facial expression, other writers have shown that outright verbal reinforcement, intentional or nonintentional misrecording of data, or misjudgment of subject responses also frequently occur.

Psychometric and Clinical Experiments

The study of expectancy effects in clinical and psychometric situations is complicated by the unreliability of psychological testing and the diagnostic process. The reliability of psychiatric and psychological judgment has been recognized as a problem in its own right for many years. Disagreements among psychiatrists, for example, have been the butt of jokes and the subject of many popular press articles designed to produce reactions ranging from amusement to indignation. The professional literature, too, has produced evidence of this problem. Kessel and Shepherd (1962), for example,

noted that a clinical judgment may be more revealing of the clinician than the child. Stuart (1970) suggested four factors that contribute to the low reliability associated with diagnostic inquiry. First, differences in individual clinicians may influence them to ask different questions and therefore elicit different behaviors from their subjects. These clinician differences also cause differences in the interpretation of behavior and differences in reasoning and observation. Secondly, irrelevant characteristics of the child such as dress, social class, attitude, and race can sometimes influence clinical judgment. Also, the category system of diagnosis is faulty, according to Stuart. Differences between the categories are based upon assumptions and inferences rather than direct observations of behavior. It is rare to find actual examples of the pathological conditions found in textbook descriptions. Finally, low diagnostic reliability may result from the context in which judgments are made. The child may behave differently in an interview or test situation than with his parents, classmates, or siblings. Thus different judgments could be made depending upon where and when the child was observed.

Diagnostic unreliability has also been studied extensively with adult patients. Ash's (1949) sample of clinicians, for example, agreed less than 50% even with regard to what major diagnostic category should be used

on a given case. His findings have been supported by many other researchers (Goldfarb, 1959; Grosz & Grossman, 1964, 1968; Katz et al., 1969; Mehlman, 1959).

Reliability studies with both psychiatrists and psychologists, then, tend to show that certain internal biases on the part of individual clinicians may be responsible for some of their differences in diagnosis. The diagnostic situation may be somewhat similar to a projective personality test. The more ambiguous the situation, the more it contains projective aspects for the clinician and, perhaps, the less it contains true information regarding the patient.

Intelligence Test Expectancy Studies with Examiner-Examinee Interaction

The administration and scoring of intelligence tests has provided a medium for the largest number of experiments in the clinical and psychometric expectancy literature. The results of these experiments are mixed and appear to depend somewhat upon whether both administration and scoring are studied or if scoring only is the task of the experiment.

Larrabee and Kleinsasser (1967) in a widely quoted, but unpublished study, found significant expectancy effects on the verbal section of the Wechsler Intelligence Scale for Children (WISC). The authors provided fictitious scores to each of five examiners who

then administered a full WISC to 12 6th grade children of average intelligence. The fictitious scores were arranged so that the examiners expected either above average or below average performance from their subjects. Each subject was tested by two examiners giving either the odd or even numbered items of the test. On the average, subjects whose performance was expected to be superior obtained full scale IQs 7.5 points higher than subjects who were thought to be below average. Differences were found primarily in the verbal section of the WISC with high expectancy subjects receiving an average of over 10 IQ points more than low expectancy subjects while the performance section differences were less than three points.

Hersh (1971) used 28 graduate school students in a study designed to test the effects of positive or negative teacher referral reports upon the Stanford-Binet and Peabody Picture Vocabulary Test scores of headstart children. He found that the positive referral reports influenced his testers to give significantly higher IQs to their subjects so designated than to those children who received negative referrals. In addition, positively referred children were rated more favorably on the Stanford-Binet's "Factors Affecting Test Performance." The testers also recommended retention and special class placement more frequently for negatively referred children.



Schroeder and Kleinsasser (1972), in an experimental format similar to the Larrabee and Kleinsasser (1967) study, reported significant WISC verbal IQ differences when their 18 graduate school examiners were given positive or negative expectancy through bogus pre-test California Tests of Mental Maturity scores (dull-normal or superior). Schroeder and Kleinsasser also commented that since their experimental WISC protocols contained few scoring errors, the effects of bias seemed to have occurred before scoring. They noted that the Information, Similarities, and Vocabulary subtests of the WISC were most susceptible to bias.

In perhaps one of the first studies in expectancy using intelligence test data, Wartenberg-Ekren (1962) was unable to show significant examiner expectancy effect on the Block Design subtest of the Wechsler Adult Intelligence Scale. This failure occurred despite the fact that the examinees reported differential treatment during testing which was related to the examiner expectancy.

Gillingham's (1970) doctoral dissertation in this area did not show significant expectancy effects. Gillingham induced positive or negative expectancy in his eight graduate student examiners by exposing them to a rating sheet containing fictitious California Test of Mental Maturity and school achievement ratings together with a "predicted WISC score." Each examiner administered

a full-scale WISC to eight junior high students of average intelligence. Gillingham suspected that his failure to produce expectancy effects may have resulted from a failure of his examiners to retain the expectancy induction.

In order to determine whether the referral problem has a biasing effect upon intelligence examiners, Saunders and Vitro (1971) asked each of six graduate school examiners to administer the Stanford-Binet Intelligence Scale to 10 children. Five children were referred for possible entry into a gifted educational program and the other five were referred because of possible retardation. The authors found no significant differences between the two groups and concluded that clinical cognitive assessment is not as likely to be influenced by examiner bias, as may be possible in an experimental or nonclinical situation when the examiners do not feel as much of an ethical responsibility for their testing.

Harry Dangel (1972) also reported a study of expectancy effects using a population of 54 mentally retarded children as examinees. Three graduate student examiners administered the WISC to these children under three different expectancy conditions. Bias was induced by fictitious existing Stanford-Binet scores of 84-93 and favorable classroom comments (positive bias condition), or Stanford-Binet scores of 55-64 and

unfavorable classroom comments (negative bias condition). A control group was obtained by exposing the examiners to the child's actual pre-test Stanford-Binet score and classroom comments indicating average behavior. Dangel found that the WISC scores did not differ significantly across treatment conditions.

From the foregoing review it is impossible to reach a conclusion regarding bias effects in intelligence testing. In all of these studies, the examiner interacted with his subjects and thus the presence of bias effects may have depended upon the examiner's ability to influence his subjects, quite apart from the effects this bias may or may not have had upon his data interpretation. A study by Lasky, Felice, Moyer, Buddington, and Elliot (1973) seems to indicate that the influence of the examiner's expectations upon his subjects' test behavior can be quite substantial. Lasky et al. used administration and scoring of the Peabody Picture Vocabulary Test as the dependent variable in this study with three examiners. Two pre-test expectancy inductions were presented, one giving an inflated picture of the subject's abilities and the other giving a true report of his abilities. Examiner-examinee interaction was controlled through two reinforcement procedures. For some subjects, the examiner was instructed to use candy as a reinforcer for correct responses, for others he was

instructed to use the standard test reinforcement. Eighty retarded mental hospital patients served as subjects. Lasky et al. found that subjects in the combined standard reinforcement-inflated pre-test score treatment condition achieved an average of 9.6 IQ points more than their standard reinforcement-true pre-test score counterparts. However, when candy reinforcement was used for correct responses, no expectancy effect was observed. Apparently the primary reinforcement of candy was able to suppress or mask the negative effects of examiner expectancy that would ordinarily occur under normal testing conditions.

Intelligence Test Expectancy Studies Without Examiner-Examinee Inter- action

Studies which do not include examiner-examinee interaction provide considerably more support for the expectancy effects hypothesis than the previous studies which did include this interaction. In the following review, only one of the seven studies presented does not show significant examiner expectancy effects.

Egeland (1969) studied examiner expectancy effects in the scoring of the Comprehension, Similarities, and Vocabulary subtests of the Wechsler Intelligence Scale for Children (WISC). Forty-six graduate students were asked to "help score some short forms of the WISC" (verbal section). They were randomly assigned to three treatment conditions: positive, negative, and neutral.

Expectancy induction was accomplished by giving each scorer a referral form consisting of a reason for testing (consideration of the subject for either accelerated classes or classes for the slow learner), falsified previous IQ (130 or 80), falsified raw scores for the WISC Information and Arithmetic subtests, and false standardized achievement test scores which were either well above or below average. In the neutral condition, test scores received no referral information or WISC subtest raw scores. The results showed differences between the scoring groups on the Comprehension and Similarities subtests ($p < .01$), but not on the Vocabulary subtest. A comparison of group means revealed that significant differences were achieved between the above average and below average expectancy groups, but not between the neutral (no information) and below average groups. Thus the below average expectancy group apparently did not suffer from negative expectation so much as the above average expectancy group benefited from halo effect.

William Simon (1969) trained 72 introductory psychology students to score some 20 WISC vocabulary items previously selected from the WISC protocols of children having IQs ranging from 90 to 110. Expectancy was induced by suggesting that the child tested was reading far above or below his age level and that the



score would be used to help determine whether to hold back or skip ahead a grade in school. The group means, analyzed by t test, were shown to be significantly ($p < .05$) different.

Sattler, Hillix, and Neher (1970) have published results from two experiments showing examiner expectancy effects. In the first experiment, 15 graduate students scored WISC Vocabulary and Comprehension subtests under bright, dull, and unspecified expectancy conditions. Expectancy was induced by supplying the scorers with bogus scaled scores for the remaining WISC subtests (6-8 for the dull condition and 12-14 for the bright condition). Analysis of the data showed that the examiners' scoring of the Vocabulary, but not the Comprehension subtest, was affected by expectancy. The second experiment was devised to present an experimental situation more like an actual examination setting. In this experiment eight graduate students who had the experience of having administered, scored, and written reports for at least 26 tests were asked to score the Comprehension, Similarities, and Vocabulary subtests from the Wechsler Adult Intelligence Scale. Their stimulus was audio-recordings of read item response scripts designed to produce verbal IQs of either 130 or 90. Embedded within these response scripts were four to five ambiguous responses for each subtest. Only these



15 ambiguous response scores were analyzed for differences between the average and superior expectancy conditions. The results showed significant halo effect in that the superior examinees received more credit than the average examinees for the same ambiguous responses.

In an exceptionally well-designed set of two experiments, Sattler and Winget (1970) were able to simulate actual testing conditions, control for examiner influence upon examinees and, at the same time, study differences in examiner behavior as a result of expectancy set. Eight graduate and former graduate students administered the Verbal Scale subtests of the Wechsler Adult Intelligence Scale to 18 paid female accomplices who had memorized response scripts designed to yield verbal IQs of either 130 (superior) or 96 (average). As in the Sattler, Hillix, and Neher (1970) study, each of these scripts contained 12 identical "ambiguous" or difficult to score responses. Expectancy was induced through superior and average referral reports and through performance on items other than the 12 ambiguous responses. The testing sessions were tape-recorded to check for differences in examiner behavior. Analysis of the data produced by these experiments showed that superior examinees received more credit for ambiguous responses than average examinees ($p < .01$), that the halo effect was operating on all three subtests, and that only those examiners

who received superior referral reports gave their superior subjects significantly more credit than their average subjects ($p < .01$). Examiners did not differ in the amount of probing questions they asked of superior or average examinees. The results of both experiments show that the effect of referral reports alone did not significantly alter test scoring, but the combination of superior referral reports and superior intellectual performance on filler items did.

Expectancy effects in a foreign situation are reported by Babad, Mann, and Mar-Hayim (1975). Eighteen Hebrew University of Jerusalem graduate students scored the Wechsler Intelligence Scale for Children record of an Israeli fifth grader. Half of the scorers were given a cover sheet portraying the student as underachieving and disadvantaged while the other half were led to expect a high-achieving, upper-middle class child. These expectancies produced scoring differences in the expected direction for the verbal subtests, but significance was obtained only on the Comprehension subtest ($p < .03$).

In an effort to determine the effects of expectancy in experienced examiners, Jacobs and DeGraaf (1973) showed 32 practicing psychologists a video-taped administration of an intelligence test given to a child. Before viewing the tape they were exposed to case histories suggesting that the child was either bright or

dull. An analysis of their judgments concerning the child's performance indicated significant expectancy effect produced by the case histories. Experience in psychological testing, apparently, does not guarantee resistance to the effects of bias.

The experience variable was a major factor of interest in a study by Auffrey and Robertson (1972) which did not show significant expectancy effects. The authors used 36 examiners at three levels of experience as subjects. "Experts" were so classified if they had at least two years of professional experience and at least 50 administrations of each the Wechsler Intelligence Scale for Children (WISC) and Wechsler Adult Intelligence Scale (WAIS) to their credit. "Interns" were graduate students who had completed Wechsler test training and had previously administered between 10 and 50 of either the WISC or WAIS. The "novice" level of experience was represented by subjects currently enrolled in a graduate Wechsler test class and who had scored at least one WAIS and one WISC protocol at the time of the experiment. After reading case histories depicting either a bleak past with a strong tone of cultural and material impoverishment or a more optimistic background, the subjects were asked to score one WAIS and one WISC protocol previously prepared for the experiment. A control group also scored the protocols, but was not exposed to the

case history information. The results showed that different case histories, either by themselves, or in combination with varying levels of experience, did not significantly influence the scoring of the two intelligence tests. Scoring reliability was significantly different between the groups, however, as is shown by clearly higher variability in the novice group of examiners. Although not statistically significant, an inspection of WISC scoring differences by treatment condition shows higher scores were given to children with bleak pasts than to either those with optimistic pasts or control group children. Differences in this direction were not recorded with the adult tests. One could almost speculate that a slight trend toward helping the child make up for his unfortunate past may have existed with these examiners.

In summarizing the findings from the above literature, it appears that expectancy effects are considerably easier to produce in an experimental setting which does not provide for examiner-examinee interaction. As was shown in the teacher expectation literature, it is possible that the examiner in a test administration situation may revise his expectations as a result of perceiving test behavior which is in conflict with these expectations. There is also some evidence to suggest that positive expectancy or halo effect may be somewhat more powerful than negative expectancy effect. This

possibility is shown most clearly by the Egeland (1969) experiment and is supported in the general experimenter expectancy literature by the original Rosenthal and Fode (1961, 1963) person-perception experiments where positive expectancy groups showed more of the effect than negative expectancy groups. Expectancy effects also seem easiest to produce in relatively ambiguous scoring situations such as the Vocabulary, Similarities, and Comprehension subtests from the Wechsler intelligence scales. Although expectancy effect is difficult to show in a more straightforward scoring situation such as the Arithmetic and performance subtests of the WISC, the Lasky, et al. (1973) experiment reveals that the influence of the examiner's expectancy upon subjects' responses can be substantial, even in highly structured scoring and testing situations like the Peabody Picture Vocabulary Test.

Expectancy Effects in Personality Testing

The projective testing situation offers enough ambiguity to enhance expectancy effects in the examiner's scoring and interpretation of the data. The ambiguity of the situation also tends to increase the examiner's influence upon subjects who often, too, are looking for help in structuring their responses. Even the subtlest of cues is grasped by the examinee as is illustrated in an expectancy study reported by Masling (1965). In this

experiment 14 graduate student examiners administered the Rorschach Ink Blots to 28 subjects. Half of the examiners were led to expect that it would reflect more favorably upon them if they obtained more human response percepts than animal percepts from their subjects. The other seven examiners were given the opposite expectation. All examiners were carefully warned not to coach their subjects. The results indicated that examiners given the high animal response expectancy actually obtained one-third again as high an animal to human response ratio as did the examiners led to prize human responses. Tape-recordings of the testing sessions showed that there was no differential verbal reinforcement of subjects' responses. In addition, post-experiment questioning showed that the subjects were not aware of special interest in any particular type of response on the part of their examiners. In examining how expectancy was mediated in this experiment, Barber and Silver (1968, pp. 20-21) showed that the examiners could have taken the stimulus card away from the subjects after they gave a desired response. If the subject did not give a desired response, the examiner could have allowed the subject to continue responding until the desired response occurred. Posture, gesture, and facial expression also could have communicated the examiner's desires.

Marwit and Marcia (1967) reported very strong expectancy effects in a study with the Holtzman Ink Blots. Thirty-six advanced undergraduate examiners administered five of the Holtzman blots to 53 introductory psychology students. Some of the examiners expected a high number of responses per ink blot either because of their own hypothesis or because of expectancy induction. The remaining examiners expected few responses from their subjects. The results of this experiment showed that examiners expecting many responses because of their own hypothesis actually obtained 59% more responses from their subjects than examiners expecting fewer responses. Those examiners whose expectations of many responses were experimentally induced actually obtained 61% more responses than their counterparts expecting few responses. Although almost one-third of the examiners admitted to being aware that expectancy effects were being tested in the experiment, this awareness bore no relationship to the magnitude of the expectancy effect. In a post-experimental analysis of examiner-examinee interaction, no relationship was found between the number of verbal inquiries made of examinee responses and the actual number of responses obtained ($r = .07$). In fact, contrary to the theory that verbal conditioning may have occurred, examiners who obtained fewer responses actually questioned their subjects more than examiners who obtained many responses.

Conditioning of some kind must have occurred, however, since greater expectancy effects were found with ink blots shown later in the experiment. Perhaps the very act of questioning responses was aversive to the subjects in some way.

In a slightly more complicated study which combined the dependent variables from the previous two experiments, Marwit (1968) demonstrated strong expectancy effects. The examiners were 20 clinical psychology graduate students who were led to expect either many responses, a high proportion of which would be animal responses, or few responses with a high proportion of human content. These particular combinations of response frequency and content theoretically would not be expected from normal subjects. The results were in the direction of expectancy effects and were contrary to Rorschach theory. That examiners improved in their ability to influence their subjects is demonstrated by the fact that later contacted subjects showed greater expectancy effects than earlier contacted subjects.

Strauss (1968) was unable to show expectancy effects using experienced examiners. Five female examiners administered the Rorschach Ink Blots to six female undergraduates under two expectancy conditions and one control condition. Two subjects were expected to show an introversive experience balance (comfort and

stimulation is obtained primarily from inner resources), two were expected to show an extratensive experience balance (comfort and stimulation is obtained from outside sources), and two were tested without any expectation. The results were that actual tested experience balance was unrelated to the expectations of the examiner. In order to determine the success of expectancy induction in this experiment, Strauss asked the examiners to predict the obtained introversive-extratensive ratio for their subjects. On the average, the examiners' predictions were similar to their expectancy induction. This study suggests that examiner experience may be negatively related to expectancy effects and is contrary to Rosenthal's (1966, 1976) observation that more competent professional experimenters may be the ones to show the most expectancy effects.

In projective personality testing, examiner expectancy effects appear to depend somewhat upon the examiner's ability and/or wish to influence his subjects. This hypothesis is indirectly strengthened by Pfugrath's (1962) experiment using the Taylor Manifest Anxiety Scale (1953), a paper and pencil test of anxiety which does not allow for as much examiner influence as projective tests. In this experiment, nine graduate student counselors administered the Taylor to introductory psychology students. Three examiners were led to expect

high anxiety, three expected low anxiety, and the remaining three received no expectancy induction. The results showed no significant differences among the three treatment conditions.

Interview, Observational, and
Assessment of Therapy
Experiments

In more open-ended situations, the clinician must exercise a great deal of judgment regarding the normalcy of subject responses. Sometimes the same behavior can be judged to be pathological or normal depending upon the expectation or orientation of the clinician. Raffetto (1967) has reported an interesting study in which eight advanced undergraduate interviewers questioned 96 less advanced, paid students who had spent an hour in sensory deprivation. The interview consisted of both well-structured yes-no questions and more open-ended questions such as, "did you notice any particular feelings or sensations?" The interviewers who were led to expect hallucinatory experiences in their subjects actually reported 48% of their subjects above the grand mean for hallucinatory experiences. Only 6% of the subjects were reported above this grand mean by interviewers expecting few hallucinatory experiences. Female interviewers showed the greatest expectancy effects.

Expectations can have selective effects depending upon their relevance to dependent measures as is

illustrated by Beal's (1969) doctoral dissertation. The subjects in this experiment were 135 advanced graduate students in clinical psychology who all viewed the same motion picture of a purported therapy session. Prior to seeing the movie the subjects were provided with a "staff intake report" which varied in terms of diagnosis (neurotic, psychopathic, not stated) and motivation for treatment (high, low, not stated). There were, therefore, nine possible experimental conditions. Post-viewing ratings of the person depicted in the film were obtained from all subjects on dimensions such as attraction, prognosis, willingness to treat, and perceived similarity. Regular pauses in the film allowed for subjects to indicate what they would say. These responses were content analyzed for empathy and warmth. Beal's results showed that diagnosis suggestion created expectancy effects in the areas of attraction, empathy and warmth, but not in prognosis, willingness to treat or perceived similarity. Significant expectancy effects for motivation were seen in the subjects' ratings on prognosis, willingness to treat, and perceived similarity, but not on attraction, empathy, or warmth.

Kent, O'Leary, Diament, and Dietz (1974) published a study which shows that judgments concerning data can differ as a result of expectancy even though observations of actual behavior do not. Kent, et al., predicted a

decrease or no change in the level of recorded behavior as a function of "treatment" to two groups of observers. Both groups viewed the same video-tapes of child behavior in a classroom setting. The child behavior tapes were selected to actually show no change from "baseline" to "treatment" behavior. The results of this experiment showed no significant differences between observer groups on behavioral recordings, but the observers' global evaluations of treatment were significantly affected by the authors' predictions.

The difference between expectancy effects upon behavioral and global ratings is also shown in an experiment reported by Beasley and Manning (1973) in the field of speech pathology. Forty master's degree students in speech pathology were asked to evaluate a 10-minute, tape-recorded language sample using three objective measures and four subjective scales of language performance. Four levels of biased pre-information were presented: (a) speaker's age only, (b) speaker's age and the statement, "this child has delayed language," (c) negative pre-information, and (d) positive pre-information. The findings of this study did not show bias effects on the objective measures of language performance, but subjective ratings were significantly affected by the pre-information. Expectancy effects in the area of speech pathology and audiology have not

generally been found when objective measures are used as the dependent variable (Hipskind & Rintelmann, 1969; Meitus, Ringel, House, & Hotchkiss, 1973).

Labeling Bias in Clinical Studies

Szasz (1961) contended that psychiatric diagnosis is a process of labeling social behavior rather than a classification of diseases. The referent to which a given person's behavior is compared, then, becomes the social and ethical norms of society and psychiatry instead of observable disease processes as is contended by the professionals. This point of view is given impetus by a series of studies authored and co-authored by Maurice Temerlin. These studies provide some of the most blatant and hard hitting indictments of labeling in the literature.

In his first study of labeling bias, Temerlin (1968) trained a professional actor to portray a mentally healthy man. He was happy in his work, warm in his relationship with the interviewer, self-confident, secure, and happily married. He had but the most infrequent arguments with his wife and showed some normal concerns over the world situation. In other words, his script was systematically designed to be contra-indicative of mental illness. Just prior to listening to an audio-recording of this interview, 45 graduate students in clinical psychology, 25 practicing clinical psychologists, and 25 psychiatrists heard a "prestige confederate" remark that

the patient on the tape was "a very interesting man because he looks neurotic, but actually is quite psychotic." Three control groups were matched and stratified for professional identity. One group diagnosed the interview without prior suggestion and the other diagnosed it with the prior suggestion reversed (" . . . looks psychotic, but actually is quite neurotic"). Another group evaluated the interview as part of an "employment project." A fourth control group consisted of a mock sanity hearing with lay jurors. Subjects diagnosed the taped patient on a list of 10 psychoses, 10 neuroses, and 10 miscellaneous personality types, one of which was "normal or healthy personality." The "employment evaluators" responded to 10 scales within which was embedded a mental health scale with psychosis at one extreme and neurosis in the center. The results of this experiment showed differences between experimental and control groups to be significant at the .01 level. While no control group subject ever diagnosed psychosis, 60% of the psychiatrists, 28% of the clinical psychologists, and 11% of the graduate students did diagnose psychosis.

In discussing his results, Temerlin (1968) suggested that psychiatrists may have displayed a poor showing because organized medicine rewards conformity with prestige figures, because psychiatrists probably encounter more psychotics in their daily work than the other



professional groups, and because of the old implicit rule in medicine, "when in doubt, diagnose illness." An illness diagnosis, in an uncertain situation, is much less dangerous to a physician than a health diagnosis when illness is, in fact, present.

Temerlin and Trousdale (1969) expanded the data from the original Temerlin (1968) experiment. They show that this study included 156 undergraduate students and 40 law students besides the psychiatrists, clinical psychologists and clinical psychology graduate students reported by Temerlin (1968). When all of the diagnostic possibilities (psychoses, neuroses, and character disorders) are considered together, diagnoses of "mental illness" in this experiment ranged from a low of 84% among undergraduates to a high of 100% among the psychiatrists. Diagnoses of normality ranged from 0% to 16% in the experimental groups to 57% to 100% in the control groups. In addition to diagnosing the "patient" on the list of diagnoses, each subject was asked to write a brief description of the patient. They were instructed to report a description of behaviors which indicate personality characteristics and to avoid clinical inferences. An analysis of these reports showed a consistent failure to describe what was heard, rather than inferred, despite the explicit instructions to be descriptive. The subjects

appeared to leap to conclusions which were then defended rather than re-evaluated in the light of subsequent observation.

As part of Lee and Temerlin's (1970) study on the relationship of perceived social class and psychiatric diagnosis, the authors partially replicated the above experiment. Here, 30 psychiatric residents were asked to diagnose the same audio-recording as was used in the Temerlin (1968) experiment. Diagnosis was accomplished by rating the "patient" on a nine-point scale with psychosis and normality as end points. Before diagnosing they were told that "two board certified psychiatrists and a psychoanalyst thought the patient 'neurotic' (N=10) or 'psychotic' (N=10)." The control group (N=10) diagnosed without expectancy induction. Post-experimental analysis revealed that all subjects who expected the patient to be neurotic diagnosed him as such; 9 of the 10 subjects who received the suggestion that he was psychotic diagnosed psychosis; and 9 of the 10 controls diagnosed him as normal.

Critchley's (1970) doctoral study with 90 nursing students also shows powerful expectancy effects resulting from psychiatric labels. The nurses viewed a filmed interview of three children after being informed that each child was normal, obsessive-compulsive, or schizophrenic. After viewing the film, the nurses rated each child on an evaluation check list. The results indicated

that the subjects evaluated children labeled as schizophrenic or obsessive-compulsive as significantly more disturbed than children labeled normal. Interestingly, these subjects did not evaluate schizophrenic children as any more disturbed than those they thought were obsessive-compulsive.

In his attempt to explain some of the reasons for the relatively poor showing of psychiatrists as opposed to psychologists in his labeling bias experiments, Temerlin (1968) noted that psychologists tend to be more critical and skeptical than psychiatrists because of their historical origins in philosophy and current identifications as social scientists. This is illustrated by the heterogeneous divisions of the American Psychological Association and by the old joke, "wherever there are two psychologists, there are three opinions." In this vein, Langer and Abelson (1974) have published a study which examines expectancy and its relationship to the theoretical orientation of clinical therapists. The subjects were 40 clinicians associated with university departments known to be either behaviorally or psychodynamically oriented. Most of the subjects were psychologists, but about one-fourth were psychiatric residents. In preparation for the study, the authors video-taped an interview with a paid job applicant. Although the interview was unstructured, it centered around the interviewee's

feelings and experience relating to his past work. Bias was introduced by telling half of the subjects the interviewee was a "patient" and the other half he was a "job applicant." After viewing the video-tape, each subject was asked to write a brief, free-response description of the interviewee, his gestures, his attitudes, and the factors that probably explained his outlook on life. These descriptions were then quantified by having five graduate students blind rate them on a scale from one (very disturbed) to 10 (very well-adjusted). To make sure that the clinicians from the different universities did indeed hold different theoretical orientations, they were asked to label their orientation and to respond to a four-item rating scale designed to distinguish between behaviorists and psychodynamically oriented therapists.

An analysis of variance performed on the Langer and Abelson data revealed a significant difference between the expectancy (job applicant vs. patient) groups ($p < .01$). Analysis of the interaction between theoretical orientation and label also showed significance ($p < .05$). The "patient" label produced significantly more ascriptions of disturbance from the psychodynamically oriented clinicians than from their behaviorally oriented colleagues. On the other hand, no real differences were observed between the two theoretical groups when they thought the interviewee was a "job applicant."

Langer and Abelson speculated that the behaviorally oriented therapists may have resisted bias because their training actively encouraged a discounting of labels.

Summary and Implications for This Study

Expectancy effects have been studied through two basically different methods. One method has involved the induction of an expectancy followed by a measurement of the effects upon a second party such as a student in a teaching situation, a subject in an experimental situation, or an examinee in a testing situation. The effects of expectancy, then, were dependent upon the success of the expectancy induction (the degree to which it was believed or retained by the teacher, examiner, or experimenter) on the one hand, and the degree to which the biased person was able to influence his subject on the other. These factors have complicated the study of bias effects and account for many of the inconsistencies seen in this literature. In general, the teacher and experimenter expectation literature, whose studies have used this method of measurement, have not produced overwhelming proof of the "self-fulfilling prophecy." Expectancy effects, when second party measurement was used, appear to have been modified by many variables including age, sex, social status, personality characteristics, experimental task, ambiguity of task, length

of experiments, believability of bias induction, experimenter-subject acquaintance, experimenter experience, and the use of outcome versus process variables as dependent measures. The effects of these variables, however, have not been shown to be straightforward. For example, the sex and personality characteristics of the experimenter have interacted with these same variables in subjects in ways which suggest that these variables may have more to do with the experimenter's ability to influence subjects than expectancy effects per se. Since this study did not provide for interaction between school psychologists and the child they evaluated, it was not deemed important to include measures or analyses of these variables in the design.

A second method of measuring expectancy effects has involved the study of the attitudes and judgments of the biased person such as has been seen in experiments of diagnosis, test data interpretation, behavioral observations, assessment of psychotherapy, and interviews. Here, expectancy effects appear to have been much easier to demonstrate than in second party measurement studies. Greatest effects have usually been observed when persons making a judgment did not physically interact with their subjects, but merely observed them or their productions on a test.

Since this study appears to have been the first of its kind using school psychologists as subjects, it was decided to maximize the probability that expectancy effects would occur by adopting this second method of measuring the effects. It was reasoned that if expectancy effects could not be demonstrated in this study, it would be unlikely that any other study with this same population would show expectancy effects with psychologist-child interaction. If expectancy effects could be demonstrated in this study, on the other hand, an examination of the relationships among the variables measured would provide valuable information for the design of later experiments. To further maximize the probability of showing expectancy effects, this study utilized global evaluations and inferred judgments as dependent variables rather than structured and objective measures. Barber and Silver (1968) have shown that structured tasks tend to produce fewer expectancy effects than more ambiguous ones. Their observation is supported in this literature review most directly by the experiments of Beasley and Manning (1973) and Kent et al. (1974). Some of the examiner expectancy studies reviewed here also showed that the more ambiguous subtests of the Wechsler Intelligence Scales have demonstrated greater expectancy effects than the more structured subtests (Larabee & Kleinsasser, 1967; Schrader & Kleinsasser, 1972).

The studies reviewed here have shown that the experience variable can sometimes be important. Teacher expectancy studies have generally indicated that less experienced teachers tend to show greatest expectancy effects. The experimenter expectancy literature has been more equivocal on this variable and has sometimes shown greatest effects for experienced experimenters (Rosenthal, 1966, 1969a) while greatest effects have been shown for less experienced experimenters in others (Barber & Silver, 1968). Still others have shown the variable to be unimportant (Ingraham & Harrington, 1966). In testing, experience has not been found to be a critical variable (Auffrey & Robertson, 1972; Herish, 1971; Sattler & Theye, 1967). Brophy and Good (1974) have suggested that experience may be important in preventing expectancy effects since experienced teachers may be better resisters of expectancy induction and may also be better at perceiving evidence which disconfirms their expectations. On the other hand, Temerlin's (1968) study showed fewest expectancy effects with inexperienced subjects. Since this variable appears to have been important in many studies and since its effects have been quite unpredictable, this study has included experience as a variable to be examined further.

Stuart (1970), as well as other writers (Grosz & Grossman, 1964; Szasz, 1961) have reasoned that

clinicians' theoretical beliefs can influence their diagnostic judgments. The Langer and Abelson (1974) experiment provided the most direct test of this notion of all the studies reviewed here. Their finding that psychodynamically oriented clinicians were significantly more influenced by negative expectancy than their behaviorally oriented colleagues did much to prompt the inclusion of the theory variable in this study. Although several writers have speculated about the importance of clinicians' theoretical orientations in their practice, this author could find no studies, other than Langer and Abelson's (1974), which have examined this variable. Thus the inclusion of a theory variable in this study seemed to be even more important.

Several studies have reported findings which suggested that positive expectancy or halo effects may be somewhat more powerful than negative expectancy effects (Egeland, 1969; Rosenthal & Fode, 1961, 1963). Most expectancy studies have not been designed with a capability of evaluating possible differences in the relative power of these two expectations since a test of the resulting hypotheses would require both positive and negative expectancy induction as well as a control condition. For this reason, a control group was included in this study.

CHAPTER III

METHODOLOGY

The Sample

Forty-eight lower Michigan school psychologists from eight school district psychological services offices participated in this study. The eight school psychological services offices were randomly selected from those offices in the state of Michigan which employed three or more psychologists. The eight district offices were located in both metropolitan and rural areas. Four of the offices and 30 of the school psychologists were located within intermediate school districts while the other four offices and 18 school psychologists were part of local city school districts. The eight offices employed a total of 58 school psychologists, 10 of whom (17%) did not participate in the study. Post-experiment questioning revealed that eight of the 10 nonparticipating school psychologists had scheduling conflicts and thus could not attend the experimental session. Of the two remaining, one psychologist was "too busy" to participate, and the other was "not interested."

This sample of school psychologists contained a mixture of male and female subjects with from 0 to 18 years of professional experience. Although three of the subjects were interns, and thus had no professional experience, the mean experience level of the group was 6.375 years. Most of the subjects (75%) indicated that they had received their graduate training in school psychology while 23% said their graduate training had been in clinical psychology (2% had graduate training in both). When asked to rate their theoretical orientation on a continuum with the behavioral and psychodynamic schools of thought as extremes (see Appendix I), 60% of the subjects rated themselves more toward the behavioral side of the median while the others identified themselves more with the psychodynamic orientation. If subjects who used the middle scale points (5 & 6) on the above continuum could be described as being eclectic in their theoretical orientation, the distribution of the group's theoretical orientation changes to 52% behavioral, 23% eclectic, and 25% psychodynamic.

In addition to the above sample, a second study was conducted with 12 of the 17 school psychologists employed by Grand Rapids Public Schools in Western Michigan. Grand Rapids is a medium-sized, diversified industrial city whose schools educate approximately 31,000 students. The school district has extensive

special education programming and many of its programs are used by other school districts within the county. The mean experience level of this group was 5.4 years and all but one of the subjects had received graduate training in school psychology. They also rated their theoretical preferences with the following results: 58% behavioral and 42% psychodynamic, or 50% behavioral, 8% eclectic, and 42% psychodynamic.

Table 3.1 shows the subject characteristics for both of the above samples. Each school district's community type is also listed. In the case of intermediate districts, the various community types are listed with the number of local school districts within each type.

Experimental Stimuli

All school psychologists in this sample were presented with the case study of a very active, six-year-old boy in the first grade. Each psychologist was given a stapled packet of information concerning this child which contained the following items in sequence: a teacher referral report, a background information sheet, a filled out and scored Wechsler Intelligence Scale for Children-Revised protocol (1974), a Bender Visual Motor Gestalt Test protocol (Bender, 1946), a Kinetic Family Drawing (Burns & Kaufman, 1972), and a scored composite of a Wide Range Achievement Test protocol (Jastak, Bijou, & Jastak, 1965). Following their perusal of the above

Table 3.1

Sample Characteristics

School District	Sex		Community Type(s)	Mean Years Experience	Mean Theoretical Orientation	Training	
	M	F				School	Clinical
1	2	2	Urban Fringe	3.8	21.5	4	0
2 ^a	12	3	Urban Fringe (8) Town (2) Rural (4)	7.0	22.3	10	5
3 ^a	4	0	Town (3) Rural (14)	5.0	15.3	4	0
4 ^a	4	3	Urban Fringe (3)	7.1	19.9	4	2
5	3	2	Metro. Core	7.2	20.7	4	1
6	2	1	Urban Fringe	3.7	18.0	3	0
7	4	2	Metro. Core	10.5	17.5	2	4
8 ^a	3	1	Town (3) Rural (9)	1.5	10.8	2	2
Grand Rapids	9	3	Metro. Core	5.4	21.7	11	1

^aIntermediate School Districts

printed data, the psychologists viewed a video-taped 10-minute segment of the child in his classroom. Each of the above stimuli are discussed in more detail below.

Teacher Referral Report

Three different referral reports were prepared by the author, each of which constituted a biasing or control treatment condition for this study (see Appendix A). The teacher referral report was stapled to the top of the above information packet so that it would be the first information read by the psychologist. In addition to stating the child's name, age, birthdate, and grade, each referral report contained a brief description of his academic achievement, his classroom and playground behavior, and his approach to classroom work. The reason for the referral was also stated. Although all three referral reports described the same child's behavior, the teacher's choice of words and reflected attitude toward the child differed considerably on each one.

One referral report was negatively oriented to simulate one that could have been written by a teacher who was both upset and angry with this particular child. His behavior was described in negative and pathological terms such as "hyperactive, aggressive, and impulsive." He was referred for help with his problems and testing was requested to determine his eligibility for special education help.

A second, more positively oriented referral report described the same behavior, but in neutral and positive terms such as "energetic," "assertive," and "spontaneous." This teacher enjoyed having the child in class, but referred him for testing to " . . . see how I can help Scott best utilize his abilities and to determine if he has any particular needs."

The third type of referral was devoid of inferential language and described the child in strict behavioral terms. Specific examples of the behaviors described in the other referral reports were given along with their frequencies. The reason for this child's referral was the same as that described in the positive referral report above.

Background Information Sheet

The background information sheet contained a parent's statement, a series of short developmental history statements, and a few comments about the child's kindergarten experience (see Appendix B). This sheet was prepared by the author to present a child with a few problems and one who was very mildly delayed in his development. Care was taken, however, to present a history which was not unusual or abnormal.

Revised Wechsler Intelligence Scale
for Children Protocol

The author filled out and scored a WISC-R record form. The individual item responses were gleaned from the files of actual children tested by the author, and were selected so that the protocol would show some clinical evidence of borderline developmental problems in the areas of short-term memory (Digit Span), visual organization (Block Design and Object Assembly) and psychomotor speed (Coding). Sattler's (1974, p. 557) tables for "Significant Differences Between WISC-R Scaled Scores and Between IQ's" were consulted to make sure that none of the differences between scaled scores on this protocol reached statistical significance. Thus all of the scaled score differences on this protocol could easily have occurred by chance alone. The protocol was scored to portray a child with low average (WISC-R full scale IQ, 92) intellectual ability (see Appendix C).

Bender Visual-Motor Gestalt
Test Protocol

An actual Bender Visual-Motor Gestalt Test protocol was selected from the author's files. It was selected from the file of a child previously diagnosed as normal and showed development at a level reached by approximately 60-75% of children aged 6 years (Bender, 1946, p. 5). The protocol was somewhat expansive (it covers two sheets) and one of the designs showed

development of about a year less than the chronological age of the child (see Appendix D).

Kinetic Family Drawing

This test protocol copy was also originally obtained from the author's files of actual tested children and it, too, was selected for its borderline quality. The drawing depicted the child, his mother, his father, and two of his animals. The child's younger sister, mentioned in the background sheet, was missing from the drawing. The developmental level of the drawing was approximately normal for a six-year-old (Burns & Kaufman, 1972) and all of the figures appeared to be smiling and happy (see Appendix E).

Wide Range Achievement Test Protocol

A composite of the three sheets for this test was made so that the child's responses to individual items of the Reading, Spelling, and Arithmetic subtests could be seen on one page (see Appendix F). The test protocol was previously scored at a level consistent with the child's intelligence in the case of reading and spelling (1.2 grade equivalent) and slightly above this level in the case of arithmetic (1.6 grade equivalent). The test protocol showed the child to have learned the alphabet, but to be functioning at a pre-reading level.

The protocol was prepared by the author rather than taken from the case file of an actual child.

Video-Taped Behavior Observation

Psychologists were presented with a 10-minute video-tape showing the child in his classroom during a math assignment. The child seen in the tape was selected from 11 possibly overactive boys in the Grandville Public Schools, Grandville, Michigan. The boy selected was video-taped in his classroom for a full school day after permission to do so was obtained from his parents. The actual video-taping was done by the author from a corner of the room while class was in progress. The child seen on the tape was previously judged nonhyperactive by a school social worker. Moreover, the results of a battery of tests administered to him after the taping failed to reveal excessive hyperactivity, learning disability, or any other pathological condition as judged by the author and an associate school psychologist. The boy seen on the video-tape, then, was a normal, white, middle class child in a traditionally structured first grade suburban classroom. The 10-minute segment selected for this experiment did not show the driven quality of behavior seen in classical hyperactivity, but it did show a left-handed boy, with little task orientation, who was out of his seat approximately 30% of the time during an arithmetic seatwork assignment. He was not under

direct teacher instruction since she was in the front of the classroom with a reading group during this segment.

Dependent and Independent Measures

The dependent and independent measures used in this study were attached to the bottom of the information packet given to each school psychologist at the beginning of the experiment. The measures were sequentially arranged so that dependent measures were presented before independent measures. Each measure is discussed below, following the order in which it was presented during the actual experiment.

Severity of Pathology Scale

This scale was prepared by the author as a measure of the school psychologists' judgment of pathological severity for the child in question. It contained nine rating scales, seven of which related to separate areas of psychological functioning: "social functioning, inhibitory functioning, behavioral control, attentional processes, perceptual-motor development, cognitive organization, and relative academic functioning." The other two rating scales called for overall judgments in functioning and consistency of functioning (see Appendix G). The nine rating scales were actually continua with 10 points and upon which the psychologist could rate the child as functioning below average, average, or above average.

Each of the nine individual ratings for the above measure was scored as a deviation from the center score (5.5) on the continuum. In six of the rating scales ("Social adjustment, Inhibitory functioning, Behavioral control, Attentional processes, Cognitive organization, and Relative academic functioning") a deviation from the average score in either direction (below average or above average) was an indication of pathology and its deviation score was expressed in positive values and in increments of .5. For example, a rating of 3 on the Social adjustment scale was expressed as 2.5 and a rating of 7 was expressed as 1.5. In three of the individual rating scales ("Perceptual motor development, Overall consistency of functioning, and Overall functioning"), however, a rating in the above average range (6-10) was considered an indication of superior adjustment and therefore ratings which occurred within this range were given minus values. For example, a rating of 3 on the Perceptual-motor development scale was given a value of 2.5, but a rating of 7 was given a value of -1.5.

A single "severity of pathology" score was derived from the above nine individual ratings by a simple summation of the nine deviation scores. A high number was an indication that the psychologist rated the child's pathology as severe while a low number indicated an absence of pathology and, possibly, superior adjustment.

Diagnostic Classification

Parts of two diagnostic classification systems were printed on the second response sheet (see Appendix H) and each subject was asked to indicate his diagnostic choice on both systems. One system was entitled "Mandatory Special Education System" and included only those diagnostic classifications from the Michigan Mandatory Special Education Act (Public Act 198, 1971) which could in any way be associated with the child being evaluated. These diagnostic choices included "educable mentally impaired, learning disabled, and emotionally impaired." A fourth choice was placed at the top of the list entitled "not eligible for special education." Since only two classifications were of interest in this study, a diagnosis of pathology or the lack of it, a simple two-point nominal scale score was used for analysis purposes on this measure.

The other diagnostic classification list to which the subjects of this study responded was entitled "Classical System" and included several possible diagnoses which could have been rendered by a clinical psychologist working outside of schools. These diagnostic choices included "functioning within normal limits, developmental personality deviation, minimal brain dysfunction, neurotic reaction, and behavior disorder." A space was also provided for the subject to specify any other

diagnosis he wished if none of the above choices met with his satisfaction (see Appendix H). The score for this measure was derived in the same manner as that for the "Mandatory System" described above.

Degree of Confidence Scale

The subjects were asked to rate the degree of confidence they felt in making the above diagnostic judgments (see Appendix H). They rated their confidence on a scale from one (very unsure) to 10 (very confident).

Special Education Treatment Recommendation

A measure of the school psychologists' treatment recommendations was obtained by asking them to select one of eight special education treatment possibilities. The eight possibilities were derived from a similar list by Dunn (1973, p. 37) and consisted of special education interventions ranging from no special help at all to a full day placement in a special education classroom, part of which involved placement in a separate special school (see Appendix H). Since this list of treatment possibilities was hierarchically arranged with ever-increasing amounts of special education intervention, scoring on this measure was accomplished by numbering each possibility from one to eight with the higher numbers indicating more treatment involvement.

Referral for Pediatric Determination of Medication Possibilities

A measure of the school psychologists' willingness to refer this child for medication was obtained by simply asking them to respond, in yes or no fashion, to the question, "Would you refer this child to his pediatrician for determination of medication possibilities?" The score for this measure was based on a two-point nominal scale for analysis purposes (see Appendix I).

Theoretical Orientation

A measure of the subjects' theoretical preference, behavioral versus psychodynamic, was obtained by asking them to rate, on a six-point scale, their relative agreement with four statements adapted from those used by Langer and Abelson (1974, p. 7) for the same purpose. The statements presented areas of disagreement between behavioral and psychodynamic schools of thought. In addition, the subjects were asked to rate their theoretical preference on a continuum from one (strongly behavioristic) to 10 (strongly psychodynamic) (see Appendix I). A single measure for this independent variable was obtained by summing the scale points (with statement number one reversed) on all five rating continua. Thus scores of between 5 and 34 were possible with scores above the median (19.5) indicating more agreement with the psychodynamic theoretical orientation.

Experience and Training

The measure of the school psychologists' professional experience in this study was obtained by asking them to respond to the request, "Please write the number of years you have been practicing in school psychology since your graduation from training" (see Appendix I). A measure of the type of training experienced by the subjects was also obtained. Subjects simply checked whether their graduate training was primarily in school or clinical psychology (see Appendix I).

Procedures

Sampling Procedure

A list of school psychologists in the state of Michigan, arranged by school districts, was obtained from the state consultant for school psychology (Department of Education, State of Michigan). After each school district office had been assigned a number, a table of random numbers was entered and offices (school districts) were selected to be used in this study as their number appeared in the table. In April, 1976, key school psychologists within each selected school district office were then telephoned and asked if they would poll their colleagues to determine their willingness to participate in "a case study examining the diagnostic process." A follow-up call revealed answers in the affirmative for

the first nine offices contacted and tentative dates for the experimental sessions were set for May and June, 1976. A potential sample of 63 school psychologists was thus obtained in the nine Michigan school districts. One of the school districts, employing five school psychologists, was later forced to drop out of the study because of an especially heavy work load that developed at the time the experiment was scheduled to take place. The eight remaining school districts appeared to be fairly representative of the types of communities within the state. The Grand Rapids study sample was not selected randomly as above, but was arbitrarily chosen as an example of a metropolitan school district.

Experimental Procedure

Prior to each experimental session, the author compiled the above mentioned information packets and response sheets so that each packet represented a treatment condition (negative, positive, and behavioral) as indicated by its top sheet, the teacher referral report. These packets were then piled in random sequence by a throw of a die, with enough of them in the pile to assure one for each potential subject.

A room large enough to accommodate all of the psychologists within a given school district was secured, beforehand, through the efforts of the contact person in that district. The room was usually a conference room

located within the office building that housed the school psychologists' offices. Before each experimental session the author placed a Panasonic, seven-inch, reel-to-reel video-tape deck and 19-inch, black and white television monitor in the front of the room. The equipment was tested to make sure that it was in good working order and the video-taped behavior segment was threaded and pre-set to begin at the appropriate place.

After all of the participating school psychologists within that given school district had assembled, the author thanked them for their cooperation and explained that this experiment was designed to study the relative importance of several variables involved in the diagnostic process. They were told the experiment would last for approximately 45 minutes and were given the following instructions:

You will each be given a packet of information which contains the case study material of a six-year-old, first grade boy referred for psychological testing. The top sheet of each packet shows a teacher's referral which I would like you to read before looking over the other material. Besides the referral, you will find a short background information sheet which gives the parent's impression of the boy at home, some developmental history, and some comments made by his kindergarten teacher last year. Following that is a WISC-R protocol which I administered. The test is already scored and I have included all of the WISC-R record form pages so that you can look over the quality of the boy's individual responses if you wish. Following the WISC-R is a copy of the boy's Bender Gestalt Test, a Kinetic Family Drawing, and a composite of his Wide Range Achievement Test. On this composite, I tried to show all of his actual test responses on one sheet. If you have any trouble understanding that sheet, just let me know

and I'll try to show you how to interpret it. I tried to include tests from a basic battery that might be used by most school psychologists, but just in case, is there anyone who is not familiar with one of the tests I mentioned? _____ Following the test protocols, your information packet has three response sheets on which I'd like you to show your evaluation of the boy, but before you do that you will want to see a piece of video-tape I have of him in his classroom. For now, just take a look at his referral and the other data and let me know when you are ready to see the tape. I'd appreciate it too, if you would not discuss the case with each other until after we are finished and your packets have been handed in. I'm interested in your independent judgments and your packets contain slightly different information so discussion with each other would be irrelevant to your case anyway.

The information packets, each constituting one of the three treatment conditions, were then handed to the subjects in the randomized order that had previously been arranged. The subjects were observed to make sure that the teacher referral report was read before anything else. When the group had finished reading the historical and test data, the author gave the following explanation of the video-tape:

What you are about to see is about 10 continuous minutes of this boy in his classroom. I did the video-taping from a front corner of the room with a zoom lensed camera so although it may appear at times that the camera is very close to the boy, it's only the zoom lens that makes it look that way. Although he is obviously aware that I am filming, he does not know that he is the object of interest. This video-taping was done before he was seen for testing and as far as the boy is concerned, I am just taking pictures of his whole class. The portion of his behavior you are about to see occurred just after the morning recess at about 10:30. We had already been taping for about an hour and a half so the kids were pretty accustomed to the camera and me by this time. The setting is a traditionally structured, first grade, suburban classroom in my school district, Grandville.

This school is located in a middle-class neighborhood and the kids are mostly from middle-class homes. The kids are all working on different assignments and the empty seats you will see belong to students who are in the front of the room with their teacher in a reading group. Our boy is working on a math seatwork assignment and his teacher and I both agree that the behavior you will see is fairly typical of his classroom behavior in general.

Following the above explanation, the video-tape was turned on and the subjects viewed it without comment. After the tape was shown the subjects were asked to fill in their evaluations of the child on the three response sheets located at the bottom of their information packet. Questions were handled by whispering the consultation to the subjects who asked them, except when the question had to do with the intent of a response item. In this case the explanation was given to the entire group. The response phase of the experiment usually lasted for approximately 15 minutes, after which time the packets were collected and the school district name plus a subject number were recorded on each one. In most cases, the group then discussed the case and the experiment. Questioning of each subject, at this time, revealed that no one guessed the actual intent of the experiment.

Experts' Judgments

To gain an unbiased and independent opinion regarding the functioning of the child to be evaluated by the experimental subjects, the author conducted an experimental session with four highly regarded school

psychologists in the same way as explained above, but without exposing them to the teacher referral reports. These school psychologists were not aware of the purpose of the study and were only told that their expert opinions were necessary to "standardize the experimental instruments." The average experience level for this group was 12 years. Three of the four were past officers in the Michigan Association of School Psychologists and two of them had been directors in the National Association of School Psychologists.

The responses of these experts showed that three of the four diagnosed the child as "not eligible" within the "Mandatory Special Education" classification system ($\bar{x} = 1.25$, s.d. = .433). Nor did three of them recommend any special education intervention ($\bar{x} = 1.5$, s.d. = .866) or medication ($\bar{x} = 1.75$, s.d. = .433). The fourth expert diagnosed emotional impairment under the "Mandatory" classification system and recommended that the child receive special education instructional materials and consultative services to the teacher (see Appendix H). Three of the four wrote in a mildly pathological diagnosis within the "classical" diagnostic system ($\bar{x} = 1.75$, s.d. = .433). Their "severity of pathology" scores ranged from 15.5 to 20.5 with a mean of 17.5 and standard deviation of 2.121.

Hypotheses

The hypotheses tested in this experiment were concerned with the effects of biased teacher referral reports across four dependent measures.

Hypothesis 1:

School psychologists will attribute significantly greater pathology to a child when the teacher's referral of him describes his behavior in negative and pathological terms than when that referral describes his behavior in positive or behavioral terms.

Hypothesis 2:

School psychologists will diagnose a child as having a handicapping condition within the Michigan Mandatory Special Education diagnostic classification system with significantly greater frequency when the teacher's referral of the child describes his behavior in negative and pathological terms than when the referral describes his behavior in positive or behavioral terms.

Hypothesis 3:

School psychologists will place a child within a clinically pathological diagnostic category with significantly greater frequency when the teacher's referral of the child uses negative and pathological descriptive terms than when the descriptive terms are positive or behavioral.

Hypothesis 4:

School psychologists will rate a child as needing significantly more special education intervention when the teacher's referral of the child uses negative and pathological terms than when the terms are positive or behavioral.

The relationships and associations among the variables measured in this study (theoretical orientation, length of experience, confidence, severity of pathology estimate, "Classical" diagnosis, "Mandatory" diagnosis, special education treatment recommendation, and referral for medication) were examined for agreement with relationships that would be expected on the basis of previous studies. The data for these samples of school psychologists and the significance tests based upon these data cannot be regarded as a test of expectations for any real population since these samples have been treated experimentally and are therefore not representative of a real population. Thus relationships found in these samples can only be considered as suggestive of relationships for which further evidence might be sought in future studies.

The relationships of special interest in this study were those between school psychologists' length of experience and theoretical orientation, on the one hand, and their clinical judgments, such as severity of pathology estimates, "Classical" diagnoses, "Mandatory" diagnoses, and special education treatment recommendations, on the other. In particular, a positive correlation or association was expected between experience and theoretical orientation, on the one hand, and the various examples of clinical judgment, on the other,

especially in the negatively biased teacher referral condition. Thus, as was shown by Langer and Abelson (1974), school psychologists with increasing degrees of psychodynamic theoretical preference were expected to show increasingly pathological clinical judgments when they were exposed to negatively biased teacher referral reports. This relationship, however, was not to be expected when they were exposed to positive or behaviorally oriented teacher referrals.

Although past studies have not shown clearly what to expect with regard to the experience variable, this study was sufficiently similar to the Temerlin (1968) experiment to anticipate similar results. That is, increasing experience was expected to be associated with increasing pathology in the subjects' clinical judgments when they were exposed to the negatively biased teacher referral condition. It was reasoned that inexperienced school psychologists could be expected to be influenced more by actual test data than by teacher referral reports since their more recent experience with tests would predispose them to feel more comfortable with this type of data than with the teacher reports. It was also reasoned that inexperienced school psychologists might be somewhat more reticent to diagnose pathology than their more experienced colleagues since such a diagnosis may be more fearsome to them.

All of the dependent variables ("severity of pathology," "Classical" diagnosis, "Mandatory" diagnosis, special education treatment recommendation, and referral for medication) were expected to be related or associated with each other. Since all of these variables were, in one form or another, measures of pathology; logic would dictate their relationship. However, due to the unreliability of traditional diagnostic systems, the "Classical" diagnostic measure in this study was not expected to show the same degree of inter-relatedness as the other dependent variables. The "Mandatory" diagnostic measure and the special education treatment recommendation were expected to be closely associated since their agreement is required by law (Michigan Mandatory Special Education Act, 1971).

Design and Analyses

The basic design for this experiment was a three level single factor, fixed effects model with four dependent measures (Figure 3.1). The design called for random assignment of subjects to one of three fixed expectancy or treatment conditions (T_1 - T_3) within which they were nested.

A test of the four hypotheses listed in the above section was accomplished by multivariate analysis of variance and subsequent one-way univariate analyses of variance on each of the four dependent measures. These

		Dependent Measures					
		Treatments	Subjects	M ₁	M ₂	M ₃	M ₄
Design over Subjects	T ₁		S ₁ . . n=17 . S ₁₇				
	T ₂		S ₁₈ . . n=15 . S ₃₂				
	T ₃		S ₃₃ . . n=16 . S ₄₈				

T₁: Negative teacher referral condition

T₂: Positive teacher referral condition

T₃: Behaviorally oriented teacher referral condition

M₁: "Severity of pathology" measure

M₂: "Mandatory Special Education" diagnosis

M₃: "Classical System" diagnosis

M₄: Special education treatment recommendation

Figure 3.1. Experimental Design



analyses were performed with the data from both the larger random sample of school psychologists (48 subjects) and the Grand Rapids (12 subjects) replication group.

In order to examine relationships among the independent and dependent variables in this experiment, several additional analyses were carried out. School psychologists' theoretical orientation and length of professional experience (both continuous variables) were correlated, by the Pearson product moment method, with all other variables. A correlation matrix was generated separately for each treatment group, the Grand Rapids group, and for the entire randomly selected sample of 48 school psychologists with the treatment groups combined. In situations where a continuous variable was being associated with a dichotomous variable, such as was the case with the two diagnostic labeling measures, the values generated by the product moment method were equivalent to point-biserial correlation values. It should be noted that the point-biserial correlation method assumes a true dichotomy in the variables presented as such and not a dichotomy that is artificially imposed. In this case, the dichotomy between a diagnosis of normalcy versus pathology was assumed to be true since a child in actual practice will receive one of these diagnoses. The practice of diagnostic labeling is nominally scaled and is not based upon a normally distributed continuous variable.

In addition to the four dependent and two independent variables discussed above, several other variables were measured and analyzed in this study. The school psychologists' willingness to refer a child for medication and their degree of confidence in their judgments were also included in the above correlation matrices. When two dichotomous variables were associated, Fisher's Exact Tests and Chi Square Tests were performed depending upon sample sizes. Finally, the effects of training type (clinical versus school psychology) and school district upon two dependent variables (severity of pathology and special education treatment recommendation) were analyzed through separate univariate analyses of variance.

CHAPTER IV

ANALYSIS AND DISCUSSION OF RESULTS

Chapter IV will begin with a presentation of the statistical analyses for the hypotheses of this study. This section will be followed by an analysis of the relationships and associations shown among all of the variables measured. The results of the Grand Rapids study will then be reported and, finally, a summary and general discussion of the results will be presented.

Analyses for Hypotheses

The basic and preliminary statistical test for all four of the hypotheses stated in Chapter III was a multivariate analysis of variance. This analysis allowed for a test of the effects of the teacher's referral language upon four dependent measures taken together. The results showed an F-ratio of 1.816 which, with 8 and 84 degrees of freedom, could have occurred by chance with a probability of .085. The data, then, failed to support all four of the hypotheses, each associated with one of the dependent measures, at the more acceptable .05 level.

Table 4.1 shows the results of the univariate analyses of variance carried out on the same data. As would be expected from the nonsignificant multivariate analysis above, none of the univariate analyses achieved significance; however, it appears that the "severity of pathology" scale may have been the most affected of the four dependent measures by the treatments.

Table 4.1
Random Group Univariate Analysis of Variance

Variable	Hypothesis Mean Square	F-ratio	P
H ₁ "Severity of Pathology"	42.108	2.460	.096
H ₂ "Mandatory" Diagnosis	.027	.114	.892
H ₃ "Classical" Diagnosis	.475	1.992	.148
H ₄ Special Education Treatment Recommendation	3.630	1.186	.315

Degrees of freedom for hypotheses = 2
Degrees of freedom for error = 45

Table 4.2 bears out this speculation, and also shows that the amount of variance on this measure precludes any significance that could be attached to differences between the treatment condition means.

An interesting trend also appears to have occurred. Although not statistically significant, the behaviorally oriented teacher referral treatment condition has drawn higher estimates of pathology from the subjects on all

four dependent measures than the negative teacher referral. This trend was not expected since the behaviorally oriented teacher referrals were originally intended to be a control condition.

Table 4.2
Random Group Cell Means and Standard Deviations

Variable	Treatment Conditions					
	Negative		Positive		Behavioral	
	\bar{x}	s.d.	\bar{x}	s.d.	\bar{x}	s.d.
"Severity of Pathology"	12.03	(3.52)	11.57	(4.37)	14.59	(4.51)
"Mandatory" Diagnosis	1.29	(.47)	1.33	(.49)	1.38	(.50)
"Classical" Diagnosis	1.59	(.51)	1.40	(.51)	1.75	(.45)
Special Education Treatment Recommendation	2.53	(1.77)	3.47	(1.77)	3.13	(1.71)

Analyses of Relationships and Associations

Table 4.3 shows the correlations and associations between the variables of this study for the combined data of all three treatment groups (48 subjects). These correlations and associations are also shown within the negative teacher referral condition (17 subjects) on Table 4.4, within the positive teacher referral treatment condition (15 subjects) on Table 4.5, and within the behaviorally oriented teacher referral treatment condition (16 subjects)

Table 4.3

Random Group Correlations, Associations, and Significance Levels with all Treatment Groups Combined (T_1 , T_2 , and T_3 ; $n=48$)

	Experience	Theoretical Orientation	Confidence	Severity of Pathology	Special Education Treatment Recommendation
Theoretical Orientation	.370** p=.005				
Confidence	.161 p=.137	.022 p=.441			
"Severity of Pathology"	.040 p=.394	.471** p=.001	.109 p=.231		
Special Education Treatment Recommendation	-.019 p=.448	.301* p=.019	.201 p=.086	.525** p=.001	
"Classical" Diagnosis	.093 p=.265	.274* p=.030	-.138 p=.174	.308* p=.016	.326* p=.012
"Mandatory" Diagnosis	-.034 p=.410	.427** p=.001	.055 p=.355	.416** p=.002	.653** p=.001
Referral for Medication	.036 p=.403	.453** p=.001	.059 p=.344	.498** p=.001	.573** p=.001

* p < .05

** p < .01

Table 4.4
Random Group Correlations, Associations, and Significance Levels within the
Negative Treatment Condition (T₁, n=17)

	Experience	Theoretical Orientation	Confidence	Severity of Pathology	Special Education Treatment Recommendation
Theoretical Orientation	.396 p=.058				
Confidence	.225 p=.192	-.208 p=.212			
"Severity of Pathology"	-.162 p=.267	.383 p=.064	.213 p=.206		
Special Education Treatment Recommendation	-.075 p=.387	.130 p=.310	.265 p=.152	.433* p=.041	.536* p=.013
"Classical" Diagnosis	.123 p=.319	.470* p=.029	-.062 p=.407	.375 p=.069	.702** p=.001
"Mandatory" Diagnosis	-.155 p=.277	.168 p=.260	-.067 p=.400	.467* p=.029	.230 p=.122
Referral for Medication	.059 p=.411	.486* p=.024	-.123 p=.319	.500* p=.020	

* p < .05

** p < .01

Table 4.5

Random Group Correlations, Associations, and Significance Levels within the
Positive Treatment Condition (n=15)

	Experience	Theoretical Orientation	Confidence	Severity of Pathology	Special Education Treatment Recommendation
Theoretical Orientation	.493 [*] p=.031				
Confidence	-.132 p=.319	.194 p=.245			
"Severity of Pathology"	.304 p=.135	.579 [*] p=.012	.322 p=.121		
Special Education Treatment Recommendation	.074 p=.397	.270 p=.166	.450 [*] p=.046	.764 ^{**} p=.001	
"Classical" Diagnosis	.410 p=.064	.531 [*] p=.021	.058 p=.418	.471 [*] p=.038	.494 [*] p=.031
"Mandatory" Diagnosis	.188 p=.252	.785 ^{**} p=.001	.503 [*] p=.028	.626 ^{**} p=.006	.552 [*] p=.016
Referral for Medication	.188 p=.252	.127 p=.327	.403 p=.068	.626 ^{**} p=.006	.801 ^{**} p=.001

* p < .05

** p < .01

on Table 4.6. Each variable will be presented below together with a discussion of its related or associated variables when an uncorrected significance level of .01 or less was reached. The raw data frequency distributions for these variables are presented in Appendix J.

Length of Experience

The length of experience variable in the first column on Table 4.3 (combined data group) shows a significant relationship with only one other variable, theoretical orientation ($r = .37$, $p = .005$). Thus it appears that the more experienced school psychologists in this sample tended to be somewhat more psychodynamically oriented than their less experienced colleagues. This relationship was strongest in the positive teacher referral situation (Table 4.5, $r = .493$, $p = .031$) and weakest in the behaviorally oriented teacher referral situation (Table 4.6, $r = .304$, $p = .126$).

The expectation that school psychologists' length of professional experience would show greatest levels of relationship and association with the dependent variables in the negative teacher referral treatment condition was not realized in this study. An inspection of Table 4.4 shows that the highest association reached in this treatment condition was a very unimpressive point-biserial correlation of .123 between the experience variable and the "Classical" diagnosis measure.

Table 4.6

Random Group Correlations, Associations, and Significance Levels within the
Behaviorally Oriented Teacher Referral Condition (n=16)

	Experience	Theoretical Orientation	Confidence	Severity of Pathology	Special Education Treatment Recommendation
Theoretical Orientation	.304 p=.126				
Confidence	.443 [*] p=.043	.246 p=.179			
"Severity of Pathology"	.095 p=.360	.506 [*] p=.023	-.169 p=.265		
Special Education Treatment Recommendation	-.182 p=.250	.650 ^{**} p=.003	-.209 p=.219	.500 [*] p=.024	
"Classical" Diagnosis	-.203 p=.225	-.254 p=.171	-.460 [*] p=.036	-.087 p=.375	.044 p=.436
"Mandatory" Diagnosis	-.174 p=.260	.376 p=.076	-.206 p=.222	.205 p=.223	.722 ^{**} p=.001
Referral for Medication	-.096 p=.362	.711 ^{**} p=.001	.047 p=.432	.365 p=.082	.756 ^{**} p=.001

^{*} p < .05

^{**} p < .01

Theoretical Orientation

The combined data table (Table 4.3) shows that the theoretical orientation variable, in addition to its relationship with experience, was significantly related to the subjects' estimates of pathological severity for the child they evaluated ($r = .471$, $p = .001$). This relationship was strongest in the positive teacher referral group (Table 4.5, $r = .579$, $p = .012$) and weakest in the negative teacher referral treatment condition group (Table 4.4, $r = .383$, $p = .064$). Contrary to expectation, the negative and pathological language used to describe the child in the negatively biased treatment condition did not produce the strongest evidence of a relationship between these two variables.

Table 4.3 also reveals a significant association between theoretical orientation and "Mandatory" diagnosis ($r_{pbis} = .427$, $p = .001$). Here again it was the positive treatment condition that showed the strongest association between these two variables ($r_{pbis} = .785$, $p = .001$), and the negative treatment condition ($r_{pbis} = .168$, $p = .260$), the weakest association. The referral for medication variable was significantly associated with theoretical orientation as well, with a point-biserial correlation of $.453$ ($p = .001$). This association was most strongly supported in the behaviorally oriented referral situation (Table 4.6, $r_{pbis} = .711$, $p = .001$).

and was least supported in the positive referral situation (Table 4.5, $r_{pbis} = .127$, $p = .327$). The behaviorally oriented teacher referral group (Table 4.6) showed a significant relationship between theoretical orientation and special education treatment recommendation ($r = .65$, $p = .003$).

For the school psychologists in this study, then, increasing agreement with the psychodynamic theoretical orientation was related to higher estimates of pathology and was associated with more frequent pathological labeling within the Michigan "Mandatory Special Education" classification system than increasing agreement with a behavioral theoretical orientation. This was especially true when the teacher's referral of a child described him in positive terms. It appears that the psychodynamically oriented school psychologists were the ones most likely to refer an overactive child for medication, especially when the teacher's referral was expressed in behavioral terms. In addition to the psychodynamically oriented school psychologists' tendency to refer for medication, they showed an increasing tendency to use special education as a treatment when the referral of a possible learning disabled boy was expressed in behavioral terms.

Confidence

Table 4.3 does not show the confidence variable to be significantly related to any of the other variables

in this study. However, Table 4.5 shows a rather interesting trend. In the positive referral situation, a mildly positive relationship was shown between confidence and special education treatment recommendation ($r = .45$, $p = .046$). A positive association was also shown with "Mandatory" Diagnosis ($r_{pbis} = .503$, $p = .028$). Thus the psychologists in this treatment condition tended to feel somewhat confident with their "Mandatory" diagnosis of pathology and recommendations for increased amounts of special education. In the behaviorally oriented situation (Table 4.6), however, they tended not to feel confident with their "Classical" diagnosis of pathology ($r_{pbis} = .46$, $p = .036$).

"Severity of Pathology"

The combined data correlation and association matrix on Table 4.3 shows a significant relationship between the "severity of pathology" measure and the special education treatment recommendation measure ($r = .525$, $p = .001$) besides the relationship with theoretical orientation already presented. This relationship was highest in the positive (Table 4.5) treatment condition ($r = .764$, $p = .001$), but was well supported by all treatment groups. The evidence is fairly strong, then, that when estimates of pathology were high, subsequent treatment was also recommended at higher levels.

The "severity of pathology" measure was also significantly associated with "Mandatory" Diagnosis (Table 4.3, $r_{pbis} = .416$, $p = .002$). This association was also most strongly supported in the positive (Table 4.5) referral group ($r_{pbis} = .626$, $p = .006$), but did not reach significance in the behaviorally oriented referral group (Table 4.6, $r_{pbis} = .205$, $p = .223$). In the general sample the referral for medication variable was significantly associated with "severity of pathology" (Table 4.3, $r_{pbis} = .498$, $p = .001$). Again, it was the positive referral group (Table 4.5) that most strongly showed this association with a point-biserial correlation of .626 ($p = .006$). Increasing estimates of pathology, then, have shown association with increasing frequency in the referral for medication of a possible learning disabled boy.

Special Education Treatment Recommendation

Table 4.3 shows that in addition to its relationship with "severity of pathology" already discussed, the special education treatment recommendation measure was significantly associated with "Mandatory" Diagnosis ($r_{pbis} = .653$, $p = .001$). This association was strongly shown in all of the referral situations, but most significantly in the behaviorally oriented referral situation (Table 4.6, $r_{pbis} = .722$, $p = .001$) and the

negatively biased referral situation (Table 4.4, $r_{pbis} = .702, p = .001$).

The referral for medication variable, true to its previous behavior in this study, has also shown a significant association with the special education treatment recommendation variable in the combined data sample (Table 4.3, $r_{pbis} = .573, p = .001$). Although the positively biased referral situation (Table 4.5, $r_{pbis} = .801, p = .001$) and the behaviorally oriented referral situation (Table 4.6, $r_{pbis} = .756, p = .001$) showed relatively high degrees of association between these two variables, the association was very weak in the negatively biased referral situation (Table 4.4, $r_{pbis} = .23$). Increasing amounts of special education intervention were associated with increasing willingness to refer for medication except when the referral to the school psychologist was negatively and pathologically biased. In addition, increasing amounts of special education intervention recommended tended to relate to higher estimates of pathology and to be associated with diagnoses of pathology in Michigan's "Mandatory Special Education" diagnostic classification system.

"Classical" Diagnosis, "Mandatory" Diagnosis and Referral for Medication

Since the associations of the two diagnostic measures and the referral for medication measure with



the continuous variables have already been reported in previous sections, only the associations among these dichotomous variables themselves are presented here. Tests of association were used in these analyses because significance levels were desired. Fisher's Exact Tests were used in analyzing associations within the individual treatment conditions because the small sample size allowed for computations of exact probabilities within the sample. These results are presented in Appendix K-2 for the negatively biased referral group, Appendix K-3 for the positively biased referral group, and Appendix K-4 for the behaviorally oriented referral group. The data for all groups combined was analyzed by the chi square and are presented in Appendix K-1. The 95% confidence level was arbitrarily chosen as acceptable for significance.

The association between the "Classical" diagnosis and referral for medication measures was not significant. However, the referral for medication and "Mandatory" diagnosis variable were significantly associated with a corrected chi square of 5.668 ($p = .017$). Almost 48% of the psychologists who diagnosed no pathology in the "Mandatory" classification system were also unwilling to refer the child for medication. This trend was generally seen in each of the individual treatment

conditions, but only approached significance ($p = .059$) in the behaviorally oriented referral group.

The combined data sample also showed a significant association between the two diagnostic measures. The corrected chi square of association for these variables was 6.696, which, with one degree of freedom, was significant at the .01 level. Significance in the association between these two variables was also reached in the negative referral ($p = .041$) and the positive referral ($p = .047$) situations. However, the association was far from significant in the behaviorally oriented referral situation where agreement on the two diagnoses was reached by only half of the sample.

In summary, the "Classical" diagnosis and "Mandatory" diagnosis variables were in close agreement, except in the behaviorally oriented referral situation. Only the "Mandatory" diagnosis variable was significantly associated with referral for medication, especially in the behaviorally oriented referral situation.

Graduate Training in School Versus Clinical Psychology

The effects of graduate training were separately analyzed on two dependent variables, "severity of pathology" and special education treatment recommendation. The technique was a one-way fixed effects model analysis of variance with the type of graduate training as the

"treatment" variable. Since one of the subjects had training in both clinical and school psychology, his data were deleted from this analysis. Most of the subjects in this study received their graduate training in school psychology ($n = 36$). The mean "severity of pathology" score for these subjects was 12.986, as opposed to a mean score of 12.682 for the clinically trained subjects ($n = 11$). Table 4.7 shows this difference to be far from significant ($p = .833$). On the special education treatment recommendation measure, school psychology trained subjects achieved a mean of 3.167 as compared with the mean of 3.0 for the clinically trained subjects. Table 4.7 shows that the difference between these two groups on the treatment recommendation measure was also nonsignificant ($p = .892$).

Table 4.7
Analyses of Variance for Graduate Training

Source	Degrees of Freedom	Mean Square	F	P
<u>"Severity of Pathology" by Training</u>				
Main effects	1	.780	.045	.833
Error	45	17.431		
Total	46	17.069		
<u>Special Education Treatment Recommendation by Training</u>				
Main effects	1	.059	.019	.892
Error	45	3.128		
Total	46	3.061		

School District

The "severity of pathology" and special education treatment variables were also analyzed by school districts to determine whether the school psychologists differed in their responses to these two variables depending upon where they worked. Table 4.8 shows the one-way fixed effects analyses of variance for these two variables with school district as the "treatment" or "main" variable. As can be observed, both the "severity of pathology" and special education treatment recommendation dependent measures were significantly different among the various school districts with F-ratios of 3.33 and 6.17 respectively ($p < .01$). Thus it appears that the school district within which the psychologists worked made a significant difference upon their judgments of pathological severity and especially upon their recommendations for special education intervention. Table 4.8 also shows the mean scores on both dependent variables by school district. As can be observed, the average "severity of pathology" score ranged from 8.25 to 15.67 depending upon the school district. The average special education treatment recommendation score ranged from 1.67 to 5.2. While it is difficult to translate the "severity of pathology" mean scores into qualitatively understandable entities, an examination of the hierarchically arranged list of special education

500000

Table 4.8

School District Groups Analyses

Source	Degrees of Freedom	Mean Square	F	P	
<u>Analysis of Variance for the "Severity of Pathology" Dependent Measure by School District Groups</u>					
Main effects	7	44.969	3.333	.007	
Error	40	13.493			
Total	47	18.181			
<u>Analysis of Variance for the Special Education Treatment Recommendation Dependent Measure by School District Groups</u>					
Main effects	7	10.755	6.172	.001	
Error	40	1.742			
Total	47	3.085			
<u>Dependent Measures</u>					
School District	n	<u>"Severity of Pathology"</u>		<u>Special Education Treatment Recommendation</u>	
		\bar{x}	s.d.	\bar{x}	s.d.
<u>School District Group Mean Scores and Standard Deviations</u>					
1	4	12.75	1.479	3.25	.433
2	15	15.67	2.618	4.07	1.691
3	4	9.75	4.437	1.75	.829
4	7	11.21	4.199	1.57	.903
5	5	14.90	4.841	5.20	.979
6	3	12.50	2.828	2.67	1.247
7	6	10.50	3.316	1.67	.471
8	4	8.25	2.165	2.25	1.299



interventions upon which the subjects responded (Appendix H) does allow for this kind of translation. It was possible, then, for the child being evaluated to have received a range of services from special education materials only to placement within a special education resource room, depending upon his (and the psychologist's) school district.

Grand Rapids Study

The multivariate analysis of variance, testing for differences on the four dependent measures across the three fixed types of referral situations, resulted in an F-ratio of 4.540. With 8 and 12 degrees of freedom, this ratio was significant at the .0097 level. Table 4.9 shows the results of the univariate analyses carried out for each of the dependent variables. An examination of this table reveals that only one of the F-ratios for the individual dependent measures reached the .05 level of significance necessary to reject its null hypothesis. The "severity of pathology" measure reached this level ($p = .014$) and the hypothesis for this variable as stated in Chapter III was:

Hypothesis 1:

School psychologists will attribute significantly greater pathology to a child when the teacher's referral of him describes his behavior in negative and pathological terms than when that referral describes his behavior in positive or behavioral terms.

REVISED 1971

Table 4.9

Grand Rapids Group Univariate Analyses of Variance

Variable	Hypothesis Mean Squares	Univariate F-ratio	P
H ₁ "Severity of Pathology"	68.723	7.099	.014
H ₂ "Mandatory" Diagnosis	.025	.086	.918
H ₃ "Classical" Diagnosis	.125	.794	.481
H ₄ Special Education Treatment Recommendation	2.025	1.097	.375

Degrees of freedom for hypotheses = 2

Degrees of freedom for error = 45

The tests for this hypothesis were post-hoc comparisons between the negative referral treatment condition and the other two referral treatment condition means. Two tailed t-tests were used for these comparisons because, with the small sample size ($n = 12$), they were slightly more powerful than techniques using the F statistic. Table 4.10 shows the means and standard deviations for these data.

The 98% confidence interval for the difference between the negative referral and behaviorally oriented referral treatment condition means was 1.08 ± 7.80 and nonsignificant. The 98% confidence interval for the difference between the negative referral and positive referral treatment condition means was 7.43 ± 7.14 which was significant. Hypothesis 1, then, was only partially

supported by the data in this experiment. Only the difference between the negatively and positively biased teacher referral conditions reached significance on the "severity of pathology" measure.

Table 4.10

Grand Rapids Group Cell Means and Standard Deviations

Variable	Treatment Conditions					
	Negative		Positive		Behavioral	
	\bar{x}	s.d.	\bar{x}	s.d.	\bar{x}	s.d.
"Severity of Pathology"	17.33	(5.80)	9.90	(1.67)	16.25	(1.71)
"Mandatory" Diagnosis	1.67	(.58)	1.60	(.55)	1.75	(.50)
"Classical" Diagnosis	1.67	(.58)	2.00	(.00)	1.75	(.50)
Special Education Treatment Recommendation	3.33	(2.08)	4.60	(.89)	4.75	(1.25)

Tables 4.11 and 4.12 show the correlations and associations between pairs of variables for this sample. Raw data frequency distributions for these variables are also presented in Appendix J. As can be observed, the "Classical" diagnosis measure was significantly associated with the confidence measure ($r_{pbis} = .600$, $p = .020$), and the special education treatment recommendation measure ($r_{pbis} = .800$, $p = .001$). The negative association between "Classical" diagnosis and the "severity of



Table 4.11

Grand Rapids Group Correlations, Associations, and Significance Levels with
all Treatment Groups Combined (T_1 , T_2 , and T_3 ; $n=12$)

	Experience	Theoretical Orientation	Confidence	"Severity of Pathology"	Special Education Treatment Recommendation
Theoretical Orientation	.067 p=.418				
Confidence	-.169 p=.299	.502* p=.048			
"Severity of Pathology"	.304 p=.168	-.353 p=.130	-.392 p=.104		
Special Education Treatment Recommendation	-.186 p=.282	.180 p=.288	.426 p=.084	-.543* p=.034	
"Classical" Diagnosis	-.147 p=.325	.186 p=.281	.600* p=.020	-.660** p=.010	.800** p=.001
"Mandatory" Diagnosis	.357 p=.127	.338 p=.141	-.158 p=.312	-.409 p=.094	.314 p=.160
Referral for Medication	.378 p=.113	-.042 p=.449	.298 p=.173	.067 p=.418	.127 p=.347

* $p < .05$

** $p < .01$



Table 4.12

Grand Rapids Group Fisher's Exact Tests with all
Treatment Groups Combined (n=12)

	Referral for Medication		Row Total
	yes	no	
<u>"Classical" Diagnosis</u>			
Normal	n=1	n=1	n=2
	row % 50.0	row % 50.0	% = 16.7
	col % 16.7	col % 16.7	
	tot % 8.3	tot % 8.3	
Pathology	n=5	n=5	n=10
	row % 50.0	row % 50.0	% = 83.3
	col % 83.3	col % 83.3	
	tot % 41.7	tot % 41.7	
Column Total	n=6	n=6	
	% = 50.0	% = 50.0	p = .773
<u>"Mandatory" Diagnosis</u>			
"Not Eligible"	n=2	n=2	n=4
	row % 50.0	row % 50.0	% = 33.3
	col % 33.3	col % 33.3	
	tot % 16.7	tot % 16.7	
Pathology	n=4	n=4	n=8
	row % 50.0	row % 50.0	% = 66.7
	col % 66.7	col % 66.7	
	tot % 33.3	tot % 33.3	
Column Total	n=6	n=6	
	% = 50.0	% = 50.0	p = .727
<u>"Classical" Diagnosis</u>			
<u>"Mandatory" Diagnosis</u>	<u>Normal</u>	<u>Pathology</u>	<u>Row Total</u>
"Not Eligible"	n=1	n=3	n=4
	row % 25.0	row % 75.0	% = 33.3
	col % 50.0	col % 30.0	
	tot % 8.3	tot % 25.0	
Pathology	n=1	n=7	n=8
	row % 12.5	row % 87.5	% = 66.7
	col % 50.0	col % 70.0	
	tot % 8.3	tot % 58.3	
Column Total	n=2	n=10	
	% = 16.7	% = 83.3	p = .909



pathology" measure ($r_{pbis} = -.660$) in this group is particularly interesting since it means that increasing estimates of pathological severity were associated with a diagnosis of normality in the "Classical" diagnostic classification system. The "severity of pathology" measure was also negatively correlated with the special education treatment recommendation measure ($r = -.543$, $p = .034$). Inspection of the data revealed that only two subjects in this group diagnosed normality within the "Classical" system. These two subjects averaged a "severity of pathology" score of 20.25 as compared with a mean score of 13.88 for the entire group. These two subjects also recommended less special education intervention than any of the other members of their school psychology staff. These associations, then, rather than being a reflection of the general trend in Grand Rapids, were based primarily upon the behavior of two psychologists who were somewhat atypical of their group.

In summary, the results of this replication study showed that negatively, positively, and behaviorally oriented referrals produced significant ($p < .05$) differences among the subjects' clinical judgments concerning the child they evaluated. Only the "severity of pathology" measure appears to have been significantly affected by the referral conditions and post-hoc analyses showed significant ($p < .02$) differences between the

1880

1881

1882

negative and positive treatment condition means for this measure. However, the differences between the negative and behaviorally oriented treatment means failed to reach this level ($p > .02$). Most subjects diagnosed the boy they evaluated as pathological within the "Classical" (83.3%) and "Mandatory" (66.7%) diagnostic systems and almost all of the subjects (92%) believed he ought to have some special education help. A discussion with several of the subjects, after the experiment, revealed that many of the psychologists in the school district feel pressure from their administration to help place children in special education programs. The behavior of these school psychologists, then, may have been more a reflection of the special education philosophy within their school district than anything else.

Summary and Discussion of Results

The four hypotheses for these experiments predicted that school psychologists' evaluations of a possible learning disabled boy would be significantly affected by a negatively biased teacher referral report. Each hypothesis made this prediction in relation to one of four dependent variables measuring the subjects' estimations of pathological severity for this youngster (H_1), their choice of diagnostic label within the Michigan "Mandatory Special Education" (H_2) and



"Classical" (H_3) diagnostic classification systems, and their estimation of the amount of special education the boy would require (H_4).

The results of the study conducted with 48 practicing school psychologists from eight randomly selected school districts in lower Michigan failed to support any of the four hypotheses. The results of the same study conducted with 12 practicing school psychologists in the Grand Rapids Public School District did show significant differences among the three types of teacher referral treatment conditions. Subsequent analyses showed these differences to be significant only on the subjects' estimates of pathological severity for the child they evaluated. This finding only partially supported the hypothesis that the subjects would estimate more pathology when exposed to a negatively biased teacher referral than when exposed to either positively biased or behaviorally oriented referrals. Although the difference between the negative and positive treatment condition means was significant ($p < .02$) on this measure, the difference between the negative and behaviorally oriented treatment condition means was not ($p > .02$). Several possible reasons for the negative findings of the major experiment are advanced below.

Brophy and Good (1974) have suggested that many failures to demonstrate expectancy effects may be due



to a failure in the induction of the expectancy itself. Since expectancy, in these two experiments, was induced through a biased, written referral report, it may be that personal contact with the referring teacher would have more forcibly induced the expectancies. Or, it may be that school psychologists are influenced only by certain teachers with whom they are acquainted. Since the referring teacher was unknown to the subjects in these experiments, perhaps any chance of expectancy induction was, therefore, removed. It is also possible that school psychologists simply do not give much credence to a teacher's judgments concerning a child; and, therefore, although the referral reports were noted by the psychologist subjects, they were not influenced by them. It should be remembered, too, that all three referral reports used as treatments in these studies described the same behavior of a child and differed only in the language used to describe the behavior. No false pre-evaluation information was used to induce expectancy, only biased information was used. Most expectancy studies have used false information to induce expectancy, and therefore the induction method in these two experiments, although more realistic, may be considered to be somewhat weaker than other experiments. Further, it is possible that the positive teacher referral condition (Appendix A-2) may not have induced truly positive expectations. The



author may have failed to write a plausible "reason for referral" in this case, since such glowingly described children are rarely referred for testing in actual practice. Some past experiments have surmounted this problem by suggesting that the reason for referral was "to evaluate the child for possible placement in an accelerated class, or class for gifted children." Although this method may produce greater expectancy effects, it was not used here because it is not only false, but perhaps is too unfair to the subjects as well.

These studies may have fallen victim to the very thing they were designed to measure. All of the experimental sessions were conducted personally by the author, who is himself a school psychologist. Although considerable care was taken to avoid it, the author may have inadvertantly communicated an unconscious desire for his colleagues to resist the effects of expectancy and bias. Certainly the author would like to believe that he could resist such effects himself and perhaps this desire was projected to his colleagues as well.

The experience level of the subjects may also have been responsible for the lack of positive findings in these experiments. Although the subjects ranged from interns to 18-year veterans, even the interns were considerably more experienced in evaluation than the subjects of most other studies finding expectancy effects.

Ingraham and Harrington (1966) and Brophy and Good (1974) have each suggested that well-trained and experienced subjects may not allow their expectancies to influence their behavior as much as their less experienced counterparts.

Expectancy effects, in these studies, were measured on what Brophy and Good (1974) have termed "product" or outcome measures. Since they have shown that expectancy effects are much more likely to be observed by "process" measures or, in this case, school psychologists' interactive behavior, the fact that "process" measures were not used in these experiments may have lessened the probability of positive findings.

The sex variable may also have contributed to the failure to find expectancy effects in the main study. First, it is possible that the findings would have been different if the child to be evaluated had been a girl instead of a boy. Certainly the referred child's behavior would be considered more unusual for girls, and this novelty might have interacted with the different types of referral to produce significant expectancy effects. Secondly, the video-tape showed the referring teacher to be a woman. Some studies have shown that females may be less able to exert expectancy effects upon their subjects (Rosenthal, 1966, 1976) and thus if the referring

teacher had been male, greater expectancy effects may have been observed in these experiments.

There is some anecdotal evidence to suggest, however, that significant expectancy effects did not occur because the subjects revised their expectations in the face of disconfirming evidence. Several subjects commented after the experimental session that the referring teacher was incorrect in her description of the child, and many suggested that the teacher should be counseled regarding the proper management of active children. Several of the subjects wrote comments on the back of their response sheets such as " . . . discuss with the teacher to remove moral tone of some words--clarify what hyperactive really means," and "She needs help in aiding him to channel excess energy," and also "Inform the teacher that this is in no way a boy who could be considered for special education placement."

Despite the suggestion of several authors that less experienced experimenters and teachers show the greatest expectancy effects, it was anticipated that the most experienced subjects in these two experiments would show the greatest expectancy effects. The reasons for this departure were, first, that their suggestion was based on studies that have measured expectancy by observing its effects in a second party, a procedure not employed in these two experiments. Secondly,

Temerlin's (1968) study, which was somewhat similar to these experiments, found highly significant bias effects in the best trained and experienced clinicians and few of these effects in less experienced and trained subjects. Although a direct test of the effects of experience was not possible in these experiments, the data showed very low correlations and associations between experience and the dependent variables. These results coincide with other researchers who have not found experience to be a critical variable (Sattler & Theye, 1967; Herish, 1971; Auffrey & Robertson, 1972).

School psychologists' theoretical orientation (behavioristic versus psychodynamic) was also expected to be significantly related to their evaluation behavior on the four dependent measures when they were exposed to a negatively biased teacher referral report. The results of the main experiment failed to support this expectation. The correlations and associations between theoretical orientation and the four dependent measures failed to reach an acceptable significance level ($p > .01$) in the negatively biased referral treatment condition. Thus the rather powerful expectation effects found with psychodynamically oriented clinicians by Langer and Abelson (1974) were not duplicated in this experiment. Perhaps the subjects in this experiment were not as uniformly extreme in their theoretical preferences as Langer and Abelson's subjects.



This study did, however, find a broad, general relationship between theoretical preference and several dependent measures, but these relationships were most significant in treatment conditions other than the negative condition. It appears that increased agreement with the psychodynamic perspective was related to an increased tendency to diagnose pathology (especially within the "Mandatory" classification system), increased estimates of pathological severity, and an increased tendency to refer a child for medication. The relationship of theoretical orientation with the dependent variables appears to have been somewhat different in each type of referral situation. For example, the behaviorally oriented referral situation seems to have produced the lowest associations between theoretical orientation and diagnostic measures, but the highest with the referral for medication and special education treatment recommendation measures. It is difficult to make sense out of these differences, except that it is possible that behaviorally oriented referrals produced a need, on the part of more psychodynamically oriented psychologists, to take some action and a low need for the behaviorists to do so. Positive referral situations, on the other hand, may have produced little need for action, but more of a proclivity for pathological diagnoses on the part of

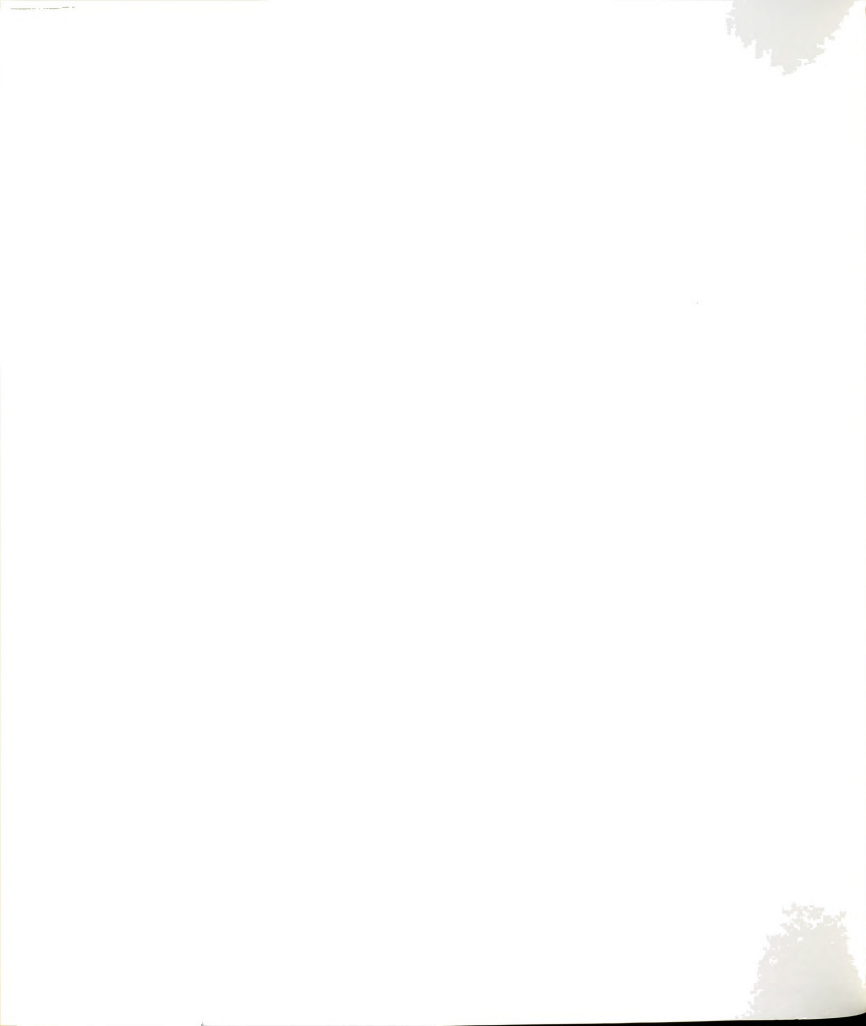
the psychodynamically oriented psychologists (and more inclination toward normal diagnoses by the behaviorists).

Despite Tolor and Brannigan's (1976) finding that their sample of school and clinical psychology trainees differed on such variables as cognitive style, perception of clients, conception of professional role, personality, sex, age, and kind and amount of previous experience; this study found no difference between the two groups, both practicing as school psychologists. Whether the school psychologists in this experiment had their graduate training in school psychology or clinical psychology did not appear to have made any difference in their estimates of pathological severity or their estimates of the amount of special education needed for the child they evaluated. It is possible that differences might have occurred if the clinically trained subjects had been working in a clinical setting. Since all of the subjects worked in schools, however, any differences that might have existed were likely attenuated by common experience.

Generally, all of the dependent measures showed significant relationships with each other. The pattern of the relationships changed somewhat in each treatment condition with the positive treatment condition showing the closest correspondence to the pattern shown when all data were combined. The "Classical" diagnosis measure

generally showed the lowest levels of association with the other dependent measures. The unreliability of the traditional diagnostic classification systems is well documented (Goldfarb, 1959; Grosz & Grossman, 1964; Szasz, 1961) and supported in this study as well. The "Mandatory" diagnostic classification system, on the other hand, was remarkably stable in its relationships with other dependent variables throughout all treatment conditions. Perhaps the reason for this stability is the fact that this system's criteria for the various diagnostic categories are somewhat more distinct than the more traditional system.

The referral for medication variable was included in these studies to gain some estimate of Michigan school psychologists' referral practices in this controversial area. The random sample of 48 school psychologists was rather evenly split in their opinion as to whether or not the boy they evaluated should be referred to his physician for determination of medication possibilities. Forty-one and seven-tenths percent indicated they would refer while 58.3% indicated they would not. They were also evenly divided on this issue within each treatment condition, except in the positive condition, where, two to one, they indicated they would not refer for medication. Thus, if the teacher's feelings about the child had any impact on the psychologist's decision whether



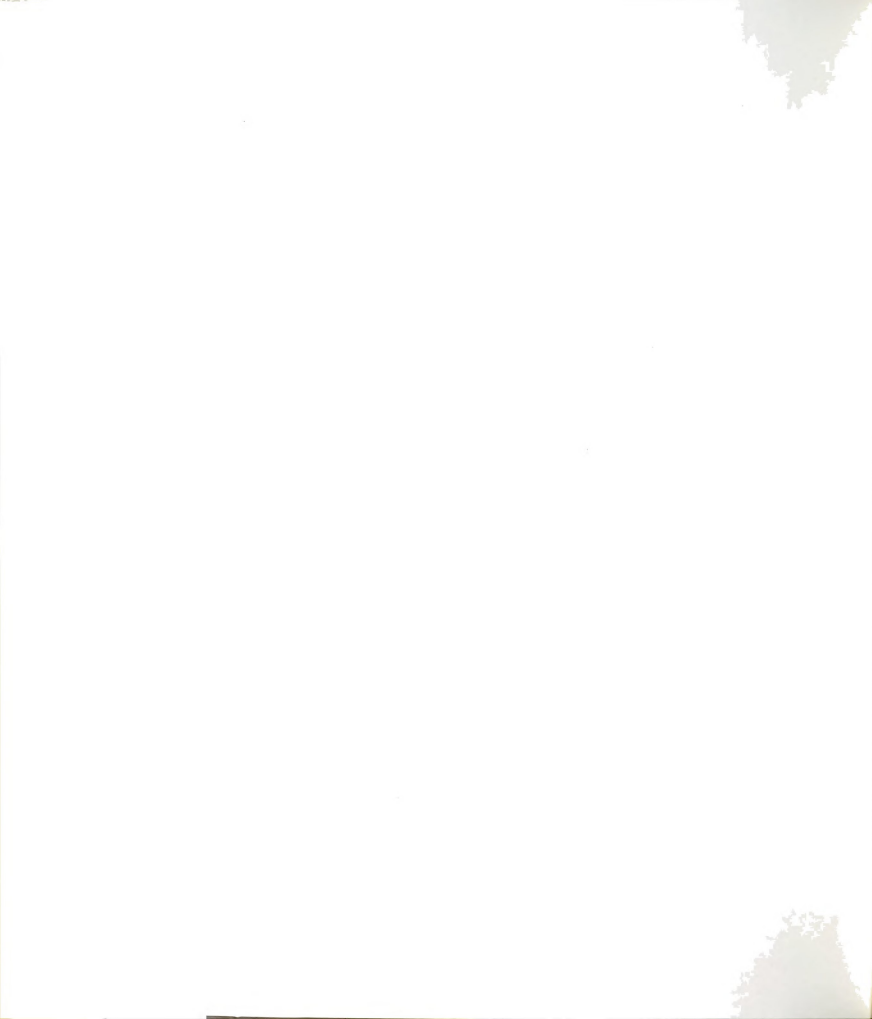
to refer for medication, it would only have been the teacher's positive feelings which helped him decide not to refer.

The results of the behaviorally oriented teacher referral report were a great surprise in these studies. Although originally intended to be a control condition, this type of referral actually showed higher estimates of pathology in the main experiment and estimates nearly as high as the negative referral condition in the Grand Rapids experiment. Although these results were not statistically significant, they were, nevertheless, in an unanticipated direction and may be indicative of a trend. If these results do reveal trends, they are difficult to understand, since the experiments did not include a control group which was not exposed to a teacher referral report. On the one hand, it is possible that the behaviorally oriented referral reports did not produce a control situation and were viewed fully as negatively by the subjects as were the negatively biased referral reports. There could have been several reasons for a negative reaction to the behaviorally oriented teacher referral. Examination of the referral report (Appendix A-3) shows that the report tends to focus primarily upon negative behavior. Perhaps the subjects were only responding to this focus. It is also possible that the frequencies of behavior reported in this referral really

do indicate pathology. The absence of published behavioral frequency norms dictated that the frequencies had to be invented entirely by the author and they may have suggested more pathology than was intended. The behavioral referral, since it did not contain evaluative or inferential statements, could also have functioned more as a screen for the subjects' projections than the other referrals. Since it is probable that most psychologists have a general bias toward pathology, this referral may have allowed some projective fruition of this general bias. On the other hand, perhaps the behaviorally oriented referral reports truly functioned as a neutral control situation. In this case, the negative referral reports were somewhat less powerful in producing expectancy effects than the positive referral reports. This interpretation would be in agreement with the results of Rosenthal and Fode (1961, 1963) and Egeland (1969) which showed relatively more powerful halo effects than negative expectancy effects. The case for a halo effect trend also gains partial support by a comparison of the "severity of pathology" mean scores for the two experimental groups with the group of experts' mean score of 17.5 on this measure. The experts, it will be remembered, were not exposed to the teacher referral reports.

Perhaps one of the major findings here is that the school psychologists in this study functioned more

similarly within each school system than between school systems. This was shown by the significant results of the analyses of variance for the eight school districts in the major experiment and by the fact that the experiment performed with a single school district (Grand Rapids) could show significant treatment effects when the same experiment performed across several school districts did not. There are several possible reasons for these significant differences between school districts. The differences could have been caused by an unequal distribution of the treatment conditions within the school districts. An examination of the data revealed that six of the eight districts received relatively equal frequencies of each treatment, but two of the districts received only two of the treatments. School district number one on Table 4.8 did not receive the positive teacher referral treatment, and district number eight did not receive the behaviorally oriented referral treatment. The absence of one of the treatments may have reduced the variance of scores within these school district groups which, in turn, could have added to the F-ratio in the analyses of variance. School district number one had particularly low standard deviations on both dependent measures. Considering the insignificant F-ratios for the univariate analyses of variance across treatment conditions (Table 4.1), this possibility seems



somewhat remote. There is no reason to assume that the treatment conditions made any real difference in subjects' responses, especially on the special education treatment recommendation variable. This variable showed the most significance on the analysis of variance across school districts (Table 4.8).

Another possible reason for the differences among the school districts is that the experimental session or the events surrounding the experiment were somehow different for the school districts. Since each experimental session was administered separately at a different time and place, over a six-week period, any number of external events could have caused psychologists within those sessions to react differently to the experiment. If differences in external events existed, however, the author was unaware of them. The possibility that some of the sessions were conducted differently also seems unlikely because each session was administered by the same experimenter (the author) and the administration was rather carefully standardized.

Three explanations for the differences among school district groups seem most plausible. One is that people who work together for a period of time may mutually influence each other to think and behave in similar ways. The literature on group processes would certainly support this notion. It is also reasonable

to suppose that the influences and pressures upon school psychologists may be somewhat different in each school district. The comments of several of the replication study subjects suggested some pressure from their special education administration to diagnose and place large numbers of children in special education programs. If this pressure truly existed, it may have accounted for the relatively high estimation of pathology and greater amount of special education treatment recommended within that group of psychologists, and perhaps some of the other groups as well. Finally, it is possible that certain of the school districts sampled contain such large numbers of deviant children that the subjects within those districts viewed the study's experimental child as needing less help than subjects who worked in districts with fewer deviant children. Thus, it is possible that local norms were influential in producing differences between the school districts.



CHAPTER V

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

Summary

Two experiments were conducted to test the effects of biased teacher referrals upon the diagnostic judgments and treatment recommendations of Michigan school psychologists. In the main experiment, 48 school psychologists from eight randomly selected school districts were randomly exposed, in school district groups, to one of three types of teacher referral reports. Each referral report described the same child, but differed in the language used for this description. The first referral report described the child in negative and pathological terms such as "hyperactive, aggressive, and impulsive"; while the second positively oriented referral report described him in terms like "energetic, assertive, and spontaneous." A third behaviorally oriented referral reported only the child's observable behavior. After reading one of the three referral reports, all of the subjects studied an identical information packet which presented the history and test protocols of an active six-year-old boy in the



first grade. All subjects also viewed the same videotaped recording, about 10 minutes in length, of the boy in his classroom. Following the viewing of the videotape, the subjects were asked to rate the child's functioning on nine continua designed to measure pathological severity. They also selected diagnostic labels from two lists taken from the Michigan "Mandatory Special Education Act" and clinical or classical diagnostic classification systems, each of which included a normal or psychologically healthy category. To gain a measure of the amount of change that they would recommend in the child's school program, the subjects were asked to select a treatment procedure from a hierarchically arranged list of special education interventions ranging from no special help at all to full-time placement in special education classrooms. Finally, the subjects were asked to indicate whether they would refer the child to his physician for determination of medication possibilities and to rate the degree of confidence they felt in making their judgments. Several independent variables were measured including the subjects' theoretical orientation (psychodynamic versus behavioral), length of professional experience, and type of graduate training (school psychology versus clinical psychology).

Four hypotheses were advanced, each of them predicting that the negatively biased teacher referral



report would influence the school psychologist subjects to overestimate the child's pathology on one of four dependent measures. The four dependent measures were: the "Severity of Pathology" scale, "Mandatory Special Education" diagnosis, clinical or "Classical" diagnosis, and the special education treatment the school psychologists decided to recommend.

The results of the main experiment failed to show significant differences among the three types of teacher referral treatment conditions on any of the four dependent measures. The same experiment, administered to a group of 12 school psychologists all employed by a medium-sized city school district, however, did result in significant ($p < .05$) differences in the subjects' estimates of pathological severity for the child they evaluated. The negative referral condition produced significantly ($p < .02$) higher estimates of pathology than the positive referral condition, but not higher than the behaviorally oriented referral condition.

An analysis of the correlations and associations among the variables measured in the main experiment revealed that as the subjects increased in experience, they tended to be somewhat more in agreement with the psychodynamic theoretical orientation ($r = .37, p = .005$). Contrary to expectation, the experience variable was not appreciably related to any of the dependent variables.

The "Severity of Pathology," "Mandatory Special Education" diagnosis, "Classical" diagnosis, special education treatment recommendation, and referral for medication variables all appeared to be in general agreement with each other. However, the "Classical" diagnosis measure agreed with the other variables at a somewhat lower level, perhaps reflecting the well-documented unreliability of traditional and clinical diagnostic classification systems. The confidence measure was not even moderately related to any of the variables measured by this experiment.

Interestingly, although the theoretical orientation variable showed moderate overall correlations and associations with the dependent measures, the pattern of these correlations and associations changed rather markedly within each type of teacher referral situation. The patterns of changes were quite complex and an understanding of why they occurred could not be determined with confidence. The positive teacher referral, rather than the negative one, showed the highest correlations and associations between theoretical orientation and the dependent measures. The direction of these relationships was that increasing levels of psychodynamic orientation were associated with increasing estimates of pathology and recommendations for special education intervention.



Whether the subjects in the main experiment received their graduate training in school or clinical psychology made no difference in their estimates of pathological severity or the amount of special education needed by the child they evaluated. Any real differences between these two groups were likely attenuated by their working together or by working outside of a clinical setting.

Surprisingly, the behaviorally oriented teacher referral treatment condition produced higher estimates of pathology on three of the four dependent measures than the negatively biased teacher referral condition in the main experiment, and nearly as high in the second experiment. Though these differences were not statistically significant, they could have been indicative of a trend in both experiments. The behaviorally oriented referral could have failed in its intended function as a control situation and could have produced as much negative expectancy as the negatively biased teacher referral. An equally, and even more plausible hypothesis, is that neither the negatively biased nor the behaviorally oriented referral reports influenced the subjects as much as did the positively biased referral.

School psychologists working together within a school district showed significantly greater agreement with each other on two dependent measures than between the various school districts. The analyses of variance



for the "severity of pathology" and special education treatment recommendation measures by school district groups in the main experiment were significant at the .007 and .001 levels respectively. The supplementary study data also showed areas of homogeneity within that group which were not seen in the combined data of the main experiment.

Conclusions

In the strictest sense, these experiments were not tests of whether expectancy effects could be demonstrated in Michigan school psychologists; for this could probably be done with enough experimental manipulation. Rather, these experiments tested whether it is reasonable to expect that Michigan school psychologists are likely to be substantially and inappropriately influenced by their primary referral agent, the teacher, when they make judgments about children. The type of referral situation of special interest was the active, somewhat developmentally immature, early elementary boy who finds school something he would rather avoid. The line between problems such as this in a normal child and those in a hyperactive child with a pathological learning problem is, many times, difficult to see in psychological evaluation. Both the symptom pattern and the diagnostic situation are extremely common in schools today. After participating in the experiment, one of the subjects

was heard to say, "I know what this boy's real name is, it's Legion." This remark is well substantiated by several incidence studies which have found these same symptoms in from one-third to almost one-half of normal boys up to nine years old (Grieger & Richards, 1976; McFarlane, Allen, & Hanzik, 1954; Werry & Quay, 1976).

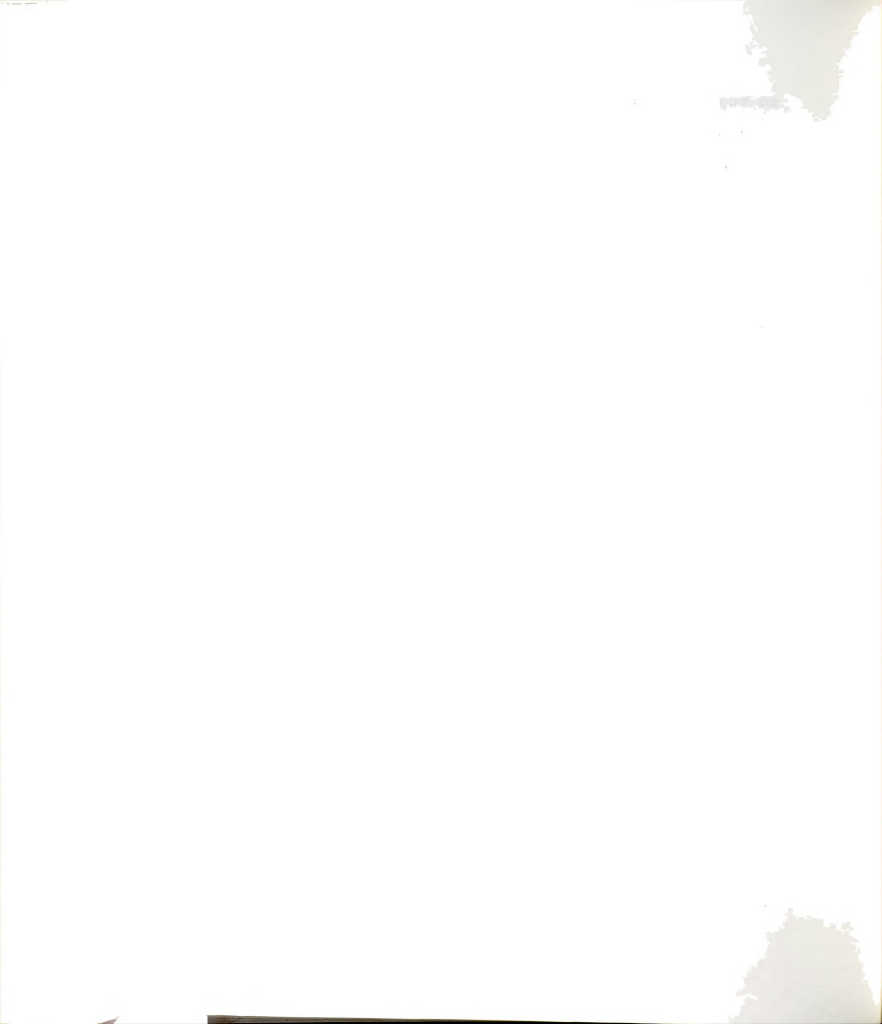
The results of the experiments indicate that, at least in situations similar to those used in the experiment, Michigan school psychologists are unlikely to be adversely affected by the bias of the referring teacher. To be sure, the replication experiment demonstrated that some effects are possible in the case of a particular school district, but it is doubtful that these effects would extend to the point of materially changing the child's educational program. The support for this conclusion comes from the fact that differences between the experimental conditions were far from significant on variables that would affect the child's life such as the "Mandatory Special Education" diagnosis, and special education treatment recommendation. Although school psychologists in certain settings may be influenced by the teacher's referral report when estimating pathology in a child, their recommendations for treatment may be influenced by other factors such as the style of remediation within a school district or the availability of special programs. In addition, there is some evidence

to suggest that placement committees like those in Michigan may make decisions for treatment quite independently of the psychologist's recommendations (Morrow, Powell, & Ely, 1976).

Why did the experiments fail to demonstrate expectancy effects? Of the many explanations possible, it appears that many of the subjects were either never influenced by the teacher referral reports, or revised their expectations when they studied the child in question. The literature of expectancy effects is so vast that experimental support could be enlisted for either of these explanations. Although the experimental task was ambiguous in these studies, the subjects were not unfamiliar or untrained with respect to what had to be done. As Ingraham and Harrington (1966, p. 460) have pointed out,

. . . Ambiguity adequately accounts for the results obtained in typical E-bias studies. Es are clearly placed in an ambiguous situation for they are naive about handling Ss and the experimental method. When they encounter Ss for the first time they have little experience to draw upon, so they must respond on cues available to them, namely, what they have been told to expect and their own varied attitudes.

Despite the realistic and commonplace nature of the evaluative problem which confronted the subjects in these studies, the fact remains that all subjects were aware that they were involved in an experimental task and not a real one. It is possible, then, that the



results would have been different in an actual practice setting. In an actual practice setting the psychologist would have actually interacted with the child which would have introduced a host of possible influencing variables. Although such interaction could have increased the power of expectancy through the psychologists' influence upon the child's behavior, some literature (Brophy & Good, 1974) suggests that the longer such interaction occurs, the less expectancy effects tend to be shown. The author's own review of the psychometric expectancy literature, too, indicates that greatest expectancy effects tend to be shown in experiments which have not provided for examiner-examinee interaction. It is submitted, therefore, that since this study did not result in significant expectancy effects when the probability for such effects was maximized through noninteraction between psychologist and client; it is unlikely that another experiment, in an actual practice setting, would produce significant effects.

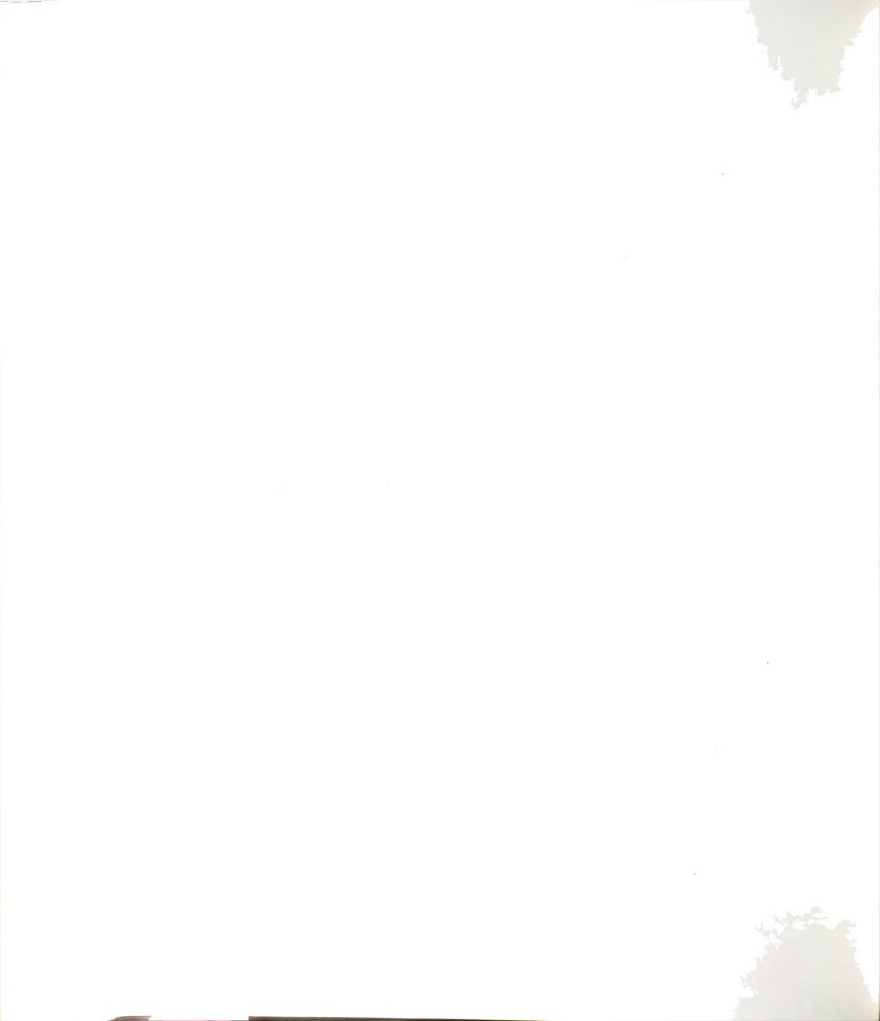
A definite conclusion regarding school district effects would not be proper here because the studies were not specifically designed to test for these effects. Hence, the statistical analyses performed were upon the same data that had shown nonsignificance in other analyses; also, the studies did not control for confounding



variables such as the effects of time and place. However, the high degree of significance reached in the main experiment ($p = .007$ and $.001$) suggests that the school district variable may be extremely influential, particularly upon the treatment recommendation activities of school psychologists. Homogeneity within a school district was also shown in the second experiment where, despite significant differences in estimates of pathological severity, the subjects tended to recommend approximately the same amount of special education intervention.

Finally, though again not part of the main experimental design and hypotheses, it is probably safe to conclude that whether school psychologists experienced their graduate training in a clinical or school psychology university program makes no difference in the pathological severity they ascribe to a child or the amount of special education intervention they recommend, in so far as the situation is like that presented in these experiments.

Three special limitations of the main study are worth mentioning as a word of caution in generalizing this study's results. First, although the sample for this study was selected randomly, the population from which it was selected included only school districts which employed three or more school psychologists. The results, therefore, cannot be generalized to school



psychologists who work alone or with a partner in independent local school districts, or to independent psychologists who contract their services to local districts. Secondly, the three types of teacher referral reports used in the experiment were treated as fixed variables and were not randomly selected from all possible or known referral reports. Thus, despite their apparent representativeness, generalization to other types of referral reports should only be done with careful attention paid to their content. Lastly, generalization should be confined to situations very similar to the type of evaluation situation presented in these experiments, that is, the evaluation of a possible learning disabled lower elementary boy.

Implications for Practice and Future Research

The results of these experiments suggest that extensive efforts on the part of school systems or school psychological service departments to reduce possible negative expectancy effects in their school psychologists are probably not necessary, at least in the context of teacher referrals. If it is felt that precautions are necessary, however, it is doubtful that a behaviorally oriented referral format, of the type used in this experiment, would be useful for this purpose. Although, as the Grand Rapids experiment showed,



it is possible for expectancy effects to occur in an individual school district; it is doubtful that these effects would substantially modify the child's education.

There is some evidence to suggest that school psychologists may be influenced by the special education philosophy within their school district or by local deviance norms within that district. At any rate, they appear to estimate pathological severity and to recommend special education intervention somewhat differently among the various school districts. Since this influence appears to make itself felt most significantly in the area of treatment recommendations, it can substantially affect the daily life and education of the child. This finding suggests that something should be done about possible misdiagnosis and unnecessary treatment of learning disability and emotional disturbance in certain school districts. Perhaps State Department of Education inservices with special education administrators and/or school psychologists would be an appropriate means of correcting this situation and standardizing the placement of children in programs. Some attention should also be given to the development of clearer, less ambiguous criteria for the diagnosis of learning disability and emotional disturbance in school children. Recognition should be given to efforts already expended in this development but further improvement

appears to be necessary. Finally, university training for school psychologists can also be addressed to this problem. For example, Bloom and Tesser's (1971) experiment found that forewarning subjects about the possibility of not being neutral reduced expectancy effects in their most influenceable subjects. Perhaps forewarning school psychologists in training about the possible influence of their school districts would be equally as beneficial. Universities would also do well to provide experiences with normal children and children with only mild problems so that trainees would be able to observe the differences between normalcy and moderate deviance.

Practicing school psychologists should be aware of the relationship that may exist between their theoretical orientation and diagnostic and recommendation practices. The psychodynamically oriented school psychologists in these experiments showed a moderate tendency to overdiagnose pathology and to overrecommend treatment. It is also possible that behaviorally oriented school psychologists may be overly influenced by positive teacher reports of a child. Further, more than a few psychologists in these studies tended to rely somewhat too heavily upon special education services for the child, even when they did not diagnose a handicapping condition under the "Mandatory" classification system. It appears that many school systems should have other

types of services available to normal children with problems, or perhaps school psychologists need more time or inclination to help classroom teachers work with these problems themselves.

One of the implications of these studies for further research is that more should be discovered about the effects of behaviorally oriented referrals and their relationship to expectancy. A replication of the main experiment reported here with the addition of more subjects and a control group receiving no referral could help to answer some of the questions about this type of referral. Does this type of referral convey neutral expectancies or do these expectancies tend to be negative? Various combinations of behavioral data together with biased teacher reports could also help to answer the question of the relative impact of behavioral data versus impressionistic or inferential teacher reports.

Future studies should also provide some method of determining whether or not the intended expectancy was actually induced in the experiment. This type of check could be accomplished through an additional control group of subjects who are asked to rate the positive, negative, or neutral qualities of the induction technique. Another method of checking the success of the induction would be to ask subjects to predict the child's adjustment on a dependent variable prior to

their seeing the observational or test data. Different post-experiment dependent variable measurements could then be compared with these predictions to determine whether earlier expectations were revised in the light of later data.

Further studies of expectancy with school psychologists would probably do well to avoid using experience as an independent variable. Its influence in these studies appears to be limited and not worth valuable space in a multivariate research design.

The studies reported here found school psychologists' theoretical orientation to be an important variable which seems to interact with dependent variables and the type of teacher referral in complex and unpredictable ways. There was some evidence that school psychologists of similar theoretical orientation may find themselves somewhat more in agreement with each other when making treatment recommendations than when diagnosing, when the referral is behaviorally oriented. Positive referrals, on the other hand, may find them more in agreement on diagnostic than treatment variables. The interaction of these variables could be studied experimentally in a multivariate design with enough subjects to include theoretical orientation as an independent variable in the analysis. A series of univariate two factor experiments could also study this interaction.

Because of these possible interactions caution should be exercised when making generalizations regarding the effects of theoretical orientation upon a single type of dependent variable.

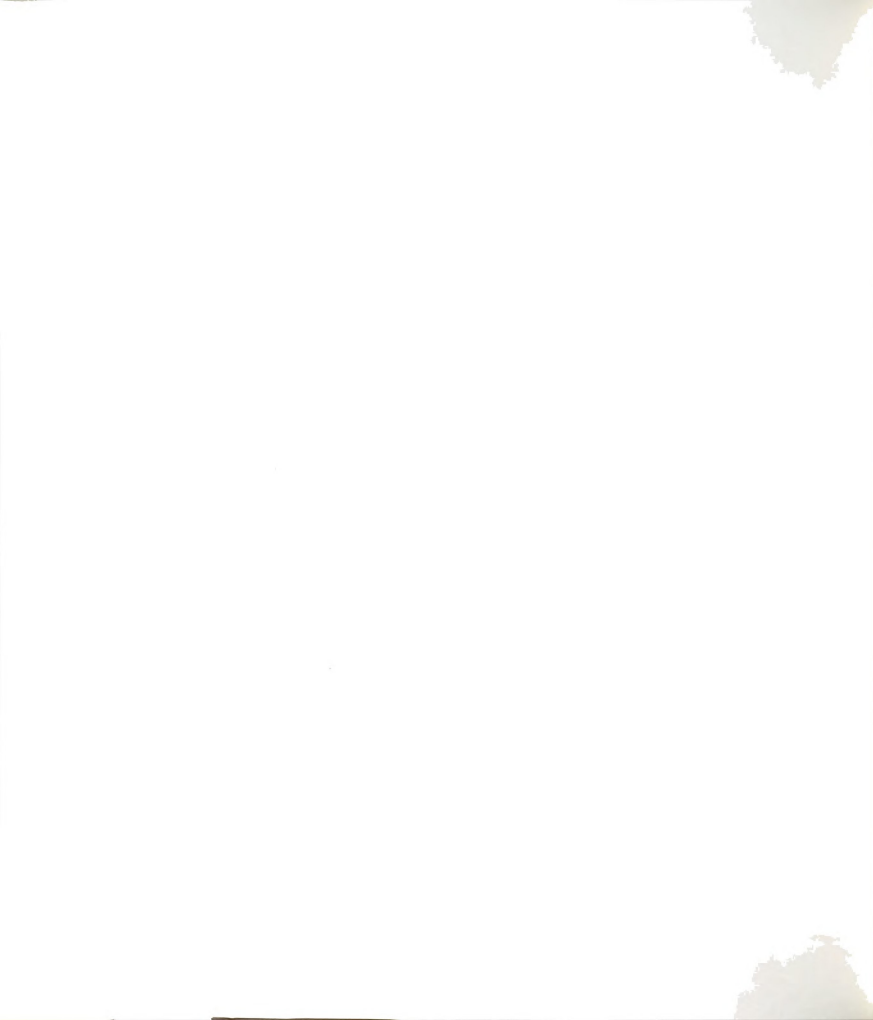
Most importantly, the influence of the employing school district upon judgments and recommendations of school psychologists should be explored further. A correlational study which attempts to identify variables within a school district which might be related to school psychologists' judgments would be an excellent beginning. For example, the number and type of special education programs proportionate to the student population in a school district could be considered an indexing variable for special education philosophy or the district's investment in special education. Various indexes of local deviance norms such as the number of Title I students or drop out rate could also be explored.

In the event that correlational studies would identify promising variables, experimental studies should be designed to test their effects. Such studies, of course, would need to control for the possibility of confounding the experimental sessions with differences in time and place. Perhaps a single large experimental session, such as during a convention or large meeting, would provide enough subjects within several school districts to adequately test hypotheses concerning

school district effects. The study could simply test for school district effects or, in a two factor design, could combine these effects with expectancy to observe possible interaction effects.



APPENDICES



APPENDIX A

TEACHER REFERRAL REPORTS

APPENDIX A
TEACHER REFERRAL REPORTS

APPENDIX A1

NEGATIVELY BIASED TEACHER REFERRAL

Name: Scott Jones
Age: 6 years, 8 months
Birthdate: August 8, 1968
Grade: 1

Brief Description of the Child:

Scott is doing poorly at the low average academic level. He is a hyperactive child in the classroom as well as on the playground. His written work is hastily done. He is aggressive and impulsive in his relationships with the other children. Scott has a concentration problem and usually can't stick with one thing for very long. His meddlesomeness and overactivity make him impossible to have in class.

Reason for Referral:

Psychological testing is requested to help Scott with his problems and to determine if he is eligible for special education help.

APPENDIX A2

POSITIVELY BIASED TEACHER REFERRAL

Name: Scott Jones
Age: 6 years, 8 months
Birthdate: August 8, 1968
Grade: 1

Brief Description of the Child:

Scott is doing acceptably well at the low average academic level. He is a very energetic youngster in the classroom as well as on the playground. His written work is quickly done. He is assertive and spontaneous in his relationships with the other children. Scott has a constant flow of ideas and usually likes a lot of variety in his work. His curiosity and spirit make him a pleasure to have in class.

Reason for Referral:

Psychological testing is requested to see how I can help Scott best utilize his abilities and to determine if he has any particular needs.

APPENDIX A3

BEHAVIORALLY ORIENTED TEACHER REFERRAL

Name: Scott Jones
Age: 6 years, 8 months
Birthdate: August 8, 1968
Grade: 1

Behavioral Description of the Child (please confine statements to objective data and/or observable behavior):

Scott is working at the 35th to 40th percentile in reading, math and spelling. During the typical hour class period Scott is out of his seat ten to fifteen times. On the playground he can be observed in at least one argument over game rules and to initiate a new game or activity (about half the time while another activity is in progress) at least once per recess. His seatwork is accurately done when assignments are short, but his accuracy decreases to 50-60% when assignments last for more than 20 minutes.

Reason for Referral:

Psychological testing is requested to help determine how best to work with Scott and to discover if he has any particular needs.



APPENDIX B

BACKGROUND INFORMATION SHEET



APPENDIX B

BACKGROUND INFORMATION

Parent's statement:

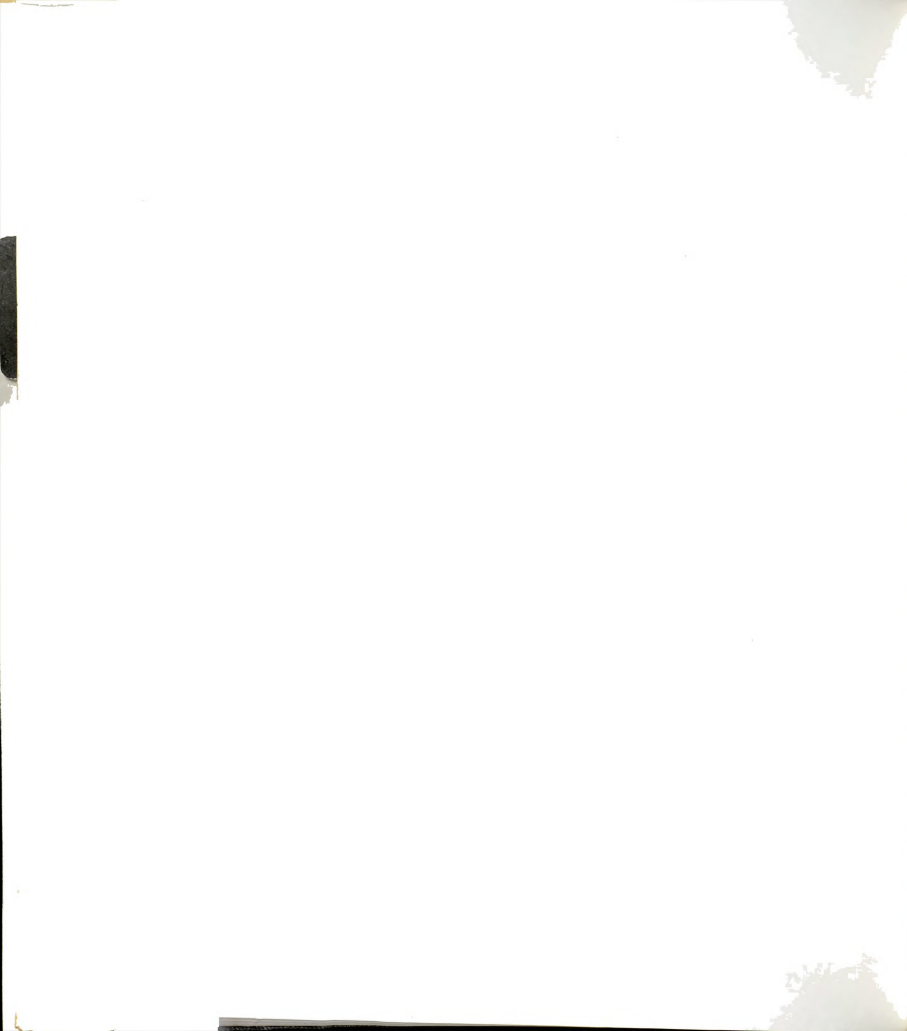
Scott's mother states that he has always been a very energetic child, constantly "getting into things" and interested in activities of his own choosing. He prefers rougher outdoor-type play to indoor activities and has never enjoyed things like coloring, painting or stories very much. He has been a fairly difficult child to discipline and is sometimes subject to temper tantrums. She is concerned that he is in the lower reading group at school and may not be doing as well as he could be. She would like as much help for him as possible.

Developmental data:

- normal pregnancy and delivery
- 7 lb. 10 oz. birth weight
- no serious childhood illnesses
- sat up without support at 7 mos.
- walked unaided at 13 mos.
- began using short sentences in speech at approximately 28 mos.
- toilet trained at 3 years, but still infrequently wets the bed.
- sleeps well, but often reluctant to go to bed
- one younger female sibling, showed some initial difficulty in adjusting to her
- high fevers, 2 days duration, at 4 mos. (infection)

School history:

- enjoyed kindergarten
- below average coloring, writing and cutting
- higher than average activity level
- preferred action-type play
- fair listening skills



APPENDIX C

WECHSLER INTELLIGENCE SCALE FOR
CHILDREN-REVISED PROTOCOL



APPENDIX C

WECHSLER INTELLIGENCE SCALE FOR CHILDREN-REVISED PROTOCOL

WISC-R

RECORD FORM

Wechsler Intelligence Scale
for Children-Revised

NAME SCOTT JONES AGE 6 SEX M

ADDRESS _____

PARENT'S NAME _____

SCHOOL _____ GRADE 1

PLACE OF TESTING _____ TESTED BY AE

REFERRED BY _____

WISC-R PROFILE

Clinicians who wish to draw a profile should first transfer the child's scaled scores to the row of boxes below. Then mark an X on the dot corresponding to the scaled score for each test, and draw a line connecting the X's.*

VERBAL TESTS							PERFORMANCE TESTS							
	Information	Similarities	Arithmetic	Vocabulary	Comprehension	Digit Span		Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes	
Scaled Score	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Scaled Score	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Scaled Score
19	19	19
18	18	18
17	17	17
16	16	16
15	15	15
14	14	14
13	13	13
12	12	12
11	11	11
10	10	10
9	9	9
8	8	8
7	7	7
6	6	6
5	5	5
4	4	4
3	3	3
2	2	2
1	1	1

*See Chapter 4 in the manual for a discussion of the significance of differences between scores on the tests.

NOTES

Outgoing, friendly
Answers quickly - tends to guess

	Year	Month	Day
Date Tested	<u>75</u>	<u>3</u>	<u>26</u>
Date of Birth	<u>68</u>	<u>7</u>	<u>8</u>
Age	<u>6</u>	<u>8</u>	<u>18</u>

	Raw Score	Scaled Score
VERBAL TESTS		
Information	<u>6</u>	<u>9</u>
Similarities	<u>6</u>	<u>9</u>
Arithmetic	<u>7</u>	<u>10</u>
Vocabulary	<u>17</u>	<u>10</u>
Comprehension	<u>11</u>	<u>11</u>
(Digit Span)	<u>(5)</u>	<u>(7)</u>
Verbal Score		<u>49</u>
PERFORMANCE TESTS		
Picture Completion	<u>12</u>	<u>10</u>
Picture Arrangement	<u>14</u>	<u>10</u>
Block Design	<u>4</u>	<u>7</u>
Object Assembly	<u>9</u>	<u>8</u>
Coding	<u>23</u>	<u>7</u>
(Mazes)	<u>()</u>	<u>()</u>
Performance Score		<u>42</u>

	Scaled Score	IQ
Verbal Score	<u>49</u>	<u>98</u>
Performance Score	<u>42</u>	<u>88</u>
Full Scale Score	<u>91</u>	<u>92</u>

*Prorated from 4 tests, if necessary.



Printed in U.S.A.

Copyright © 1971, 1974 by The Psychological Corporation.
All rights reserved. No part of this record form may be reproduced in any form of printing or by any other means, electronic or mechanical, including, but not limited to, photocopying, audio/visual recording and transmission, and portrayal or duplication in any information storage and retrieval system, without permission in writing from the publisher. See Catalog for further information.
The Psychological Corporation, New York, N.Y. 10017

74-103AS



1. INFORMATION Discontinue after 5 consecutive failures.		Score 1 or 0
1. Finger		/
2. Ears		/
3. Legs		/
4. Boil		/
5. Nickel	4	—
6. Cow		/
7. Week	5 Q 7	/
8. March	January	—
9. Bacon	a lamb	—
10. Dozen	dk	—
11. Seasons	dk	—
12. America	Geo. Wash	—
13. Stomach		
14. Sun		
15. Leap Year		
16. Bulb		
17. 1776		
18. Oil		
19. Border		
20. Ton		
21. Chile		
22. Glass		
23. Greece		
24. Tall		
25. Barometer		
26. Rust		
27. Los Angeles		
28. Hieroglyphics		
29. Darwin		
30. Turpentine		
Total		Max.=30 6

2. PICTURE COMPLETION Discontinue after 4 consecutive failures.			
	Score 1 or 0		Score 1 or 0
1. Comb	/	14. Playing Card	/
2. Woman	face	15. Girl Running	/
3. Fox	/	16. Coat	button
4. Hand	/	17. Boy	dk
5. Cat	/	18. Scissors	dk
6. Mirror	/	19. Girl	dk
7. Clock	0	20. Screw	round thing
8. Elephant	/	21. Cow	
9. Ladder	/	22. Thermometer	
10. Dresser	/	23. House	
11. Belt	dk watch	24. Telephone	
12. Man	/	25. Profile	
13. Door	/	26. Umbrella	
Total			Max.=26 12

3. SIMILARITIES Discontinue after 3 consecutive failures.		Score 1 or 0
1. Wheel—ball	wheels are bigger & ?	0
2. Candle—lamp	both are light	1
3. Shirt—hat	u can wear em	1
4. Piano—guitar	a times u can go zimmer	0
5. Apple—banana	both have peelings - stems	Score 2, 1 or 0 1
6. Beer—wine	both in a bottle & both u can drink	1
7. Cat—mouse	both can run & both have tails, they are furry	1
8. Elbow—knee	both have a hard biding right here & its hard	0
9. Telephone—radio	both have holes in em so you can hear & u can hear out of em	1
10. Pound—yard	dk	0
11. Anger—joy	both are people	0
12. Scissors—copper pan	u can both put stuff in em	0
13. Mountain—lake		
14. Liberty—justice		
15. First—last		
*16. The numbers 49 and 121		
17. Salt—water		
Total		Max.=30 6

*If the child gives a 1-point response to Item 16, say, "How else are the numbers 49 and 121 alike?"

4. PICTURE ARRANGEMENT				Discontinue after 3 consecutive failures.		
Arrangement	Time	Order	Score (Circle the appropriate score for each item.)			
Scale (SAMPLE)						
1. Fight	45"	1 13 OUT 2	0	1 OUT	2	
2. Picnic	45"	1 6 DOG 2	0	1 DOG	2	
3. Fire	45"	1 30 FIRE 2 18 FIRE	0	1 FIRE	2	
4. Plank	45"	1 35 WALK 2 20 WALK	1	1 WALK	2	
5. Burglar	45"	24 THUG	0		16-45 11-15 1-10 3 4 5	
6. Sleeper	45"	35 HRS	1		16-45 11-15 1-10 3 4 5	
7. Artist	45"	20 VAMP	0		16-45 11-15 1-10 3 4 5	
8. Lasso	45"	30 CASH	1		16-45 11-15 1-10 3 4 5	
9. Boat	60"	46 CHASE	0	2	31-40 11-20 1-10 3 4 5	
10. Gardener	60"	48 WORMS	1	2	26-40 16-25 1-15 3 4 5	
11. Bench	60"	43 BENCH	1	2	26-40 16-25 1-15 3 4 5	
12. Rain	60"	40 CLOUD	1	2	26-40 16-25 1-15 3 4 5	

* Give Sample Item First.

Max. = 48

Total 14

5. ARITHMETIC			Discontinue after 3 consecutive failures.	
Problem	Response	Score 1 or 0		
1. 30"	12	1		
*2. 30"	4	1		
*3. 30"	9	1		
4. 30"	?	0		
5. 30"	2	1		
6. 30"	4	1		
7. 30"	6	1		
8. 30"	14	1		
9. 30"	10	0		
10. 30"	12	0		
11. 30"	9	0		
12. 30"				
13. 30"				
14. 45"				
15. 45"				
16. 75"				
17. 75"				
18. 75"				
		Max. = 18		
Total		7		

* Problems 2 and 3 are given 1/2 point each if child makes error but corrects it within time limit.
† Round half-scores upward.

6. BLOCK DESIGN			Discontinue after 2 consecutive failures.			
Design	Time	Pass-Fail	Score (Circle the appropriate score for each design.)			
			1. 45"	1 17 2 13	F	①
2. 45"	1 20 2	P	0	1	②	
3. 45"	1 18 2	P	0	1	②	
4. 45"	35	F	①		21-45 16-20 11-15 1-10 4 5 6 7	
5. 75"	40	F	①		21-75 16-20 11-15 1-10 4 5 6 7	
6. 75"	75	F	①		21-75 16-20 11-15 1-10 4 5 6 7	
7. 75"			0		21-75 16-20 11-15 1-10 4 5 6 7	
8. 75"			0		26-75 21-25 16-20 1-15 4 5 6 7	
9. 120"			0		56-120 36-55 26-35 1-25 4 5 6 7	
10. 120"			0		76-120 56-75 41-55 1-40 4 5 6 7	
11. 120"			0		81-120 56-80 41-55 1-40 4 5 6 7	
			Max. = 62			
Total			4			

7. VOCABULARY Discontinue after 5 consecutive failures.		Score 2, 1, or 0
1. Knife	<i>it cuts</i>	2
2. Umbrella	<i>to use when it rains Q ?</i>	1
3. Clock	<i>to tell time with</i>	2
4. Hat	<i>u put it on so your head doesn't get cold</i>	2
5. Bicycle	<i>to ride on Q has 2 wheels</i>	2
6. Nail	<i>to pound in Q like a band</i>	2
7. Alphabet	<i>A..... Z</i>	1
8. Donkey	<i>a animal</i>	2
9. Thief	<i>a robber</i>	2
10. Join	<i>it means friends Q you're having fun</i>	0
11. Brave	<i>you're real tough Q strong</i>	0
12. Diamond	<i>its real pretty Q ?</i>	0
13. Gamble	<i>like when you're playing dice</i>	1
14. Nonsense	<i>means "well I don't know." Q ?</i>	0
15. Prevent	<i>means like that's a good picture</i>	0
16. Contagious	<i>it means Q ?</i>	0
17. Nuisance	<i>uh</i>	0
18. Fable	<i>like you eat Q ?</i>	0
19. Hazardous		
20. Migrate		
21. Stanza		
22. Seclude		
23. Mantis		
24. Espionage		
25. Belfry		
26. Rivalry		
27. Amendment		
28. Compel		
29. Affliction		
30. Obliterate		
31. Imminent		
32. Dilatory		
Total		Max.=64 17

8. OBJECT ASSEMBLY Give entire test to all children.										
Object	Time	Enter Number of Correctly Joined Cuts	Multiply by	Score (Circle the appropriate score for each item.)						
Apple (SAMPLE)										
1. Girl 120"	82	(0-6)	1	0	1	2	3	4	5	31-120 21-30 1-20 6 7 8 PERFECT ASSEMBLY
2. Horse 150"	150	(0-5)	1	0	1	2	3	4	5	36-150 21-35 16-20 1-15 5 6 7 8 PERFECT ASSEMBLY
3. Car 150"	150	(0-9)	1/2	0	1	2	3	4	5	51-150 36-50 26-35 1-25 5 6 7 8 PERFECT ASSEMBLY
4. Face 180"	180	(0-12)	1/2	0	1	2	3	4	5	76-180 51-75 36-50 1-35 6 7 8 9 PERFECT ASSEMBLY
* Round half-scores upward.										
Total									Max.=33 9	

9. COMPREHENSION Discontinue after 4 consecutive failures.		Score 2, 1, or 0
1. Cut finger	put a band-aid on it	2
2. Find wallet	call the policeman	2
*3. Smoke	call the fire engine & call the police	1
*4. Policemen	to get robbers & ?	1
5. Lose ball	find it & get him a new one	2
6. Fight	tell him to quit it	1
*7. Build house	so when a tornado comes it won't fall down	1
*8. License plates	so you know what kind of car it is & what country it's from	0
*9. Criminals	so they can't get out & steal stuff & ?	1
10. Stamps	so the mailman knows where it goes	0
11. Inspect meat	so we can have meat to eat & ?	0
*12. Charity	so they would beg & ?	0
13. Secret ballot	ah	0
*14. Paperbacks		
15. Promise		
*16. Cotton		
*17. Senators		
*If the child replies with only one idea, ask him for a second response. Rephrase the test item appropriately, saying, "Tell me another thing to do (reason why, advantage of)...."		Max.=34 11
Total		

10. CODING		
	Time	Score
A (for children under 8)	120"	23 (0-50)
B (for children 8 & older)	120"	(0-93)

CODING A Score Including Time Bonus for Perfect Performance	
Time in Seconds	Score
11-120	45
101-110	44
91-100	43
81-90	42
71-80	41
61-70	40
51-60	39
41-50	38
31-40	37
21-30	36
11-20	35

11. DIGIT SPAN (Optional) Discontinue after failure on both trials of any item. Administer both trials of each item, even if child passes first trial.					
DIGITS FORWARD					Score
	Trial 1	Pass-Fail	Trial 2	Pass-Fail	2, 1, or 0
1.	3-8-6	P	6-1-2	P	2
2.	3-4-1-7 3472	F	6-1-5-8 6158	P	1
3.	8-4-2-3-9 84329	F	5-2-1-8-6 58216	F	0
4.	3-8-9-1-7-4		7-9-6-4-8-3		
5.	5-1-7-4-2-3-8		9-8-5-2-1-6-3		
6.	1-6-4-5-9-7-6-3		2-9-7-6-3-1-5-4		
7.	5-3-8-7-1-2-4-6-9		4-2-6-9-1-7-8-3-5		
Administer DIGITS BACKWARD even if child scores 0 on DIGITS FORWARD.					Max.=14
Total Forward					
DIGITS BACKWARD					Score
	Trial 1	Pass-Fail	Trial 2	Pass-Fail	2, 1, or 0
1.	2-5	P	6-3	P	2
2.	5-7-4 457	F	2-5-9 529	F	0
3.	7-2-9-6		8-4-9-3		
4.	4-1-3-5-7		9-7-8-5-2		
5.	1-6-5-2-9-8		3-6-7-1-9-4		
6.	8-5-9-2-3-4-2		4-5-7-9-2-8-1		
7.	6-9-1-6-3-2-5-8		3-1-7-9-5-4-8-2		
Total Backward					Max.=14

Max.=28		
3	+	2 = 5
Forward	Backward	Total

8 to 9 yrs

12. MAZES (Optional) Discontinue after 2 consecutive failures.									
Maze	Maximum Errors	Errors	Score (Circle the appropriate score for each maze.)						
SAMPLE									
1. 30"	1		0	1 Error 1	0 Errors 2				
2. 30"	1		0	1 Error 1	0 Errors 2				
3. 30"	1		0	1 Error 1	0 Errors 2				
4. 30"	2		0	2 Errors 1	1 Error 2	0 Errors 3			
5. 45"	2		0	2 Errors 1	1 Error 2	0 Errors 3			
6. 60"	3		0	3 Errors 1	2 Errors 2	1 Error 3	0 Errors 4		
7. 120"	3		0	3 Errors 1	2 Errors 2	1 Error 3	0 Errors 4		
8. 120"	4		0	4 Errors 1	3 Errors 2	2 Errors 3	1 Error 4	0 Errors 5	
9. 150"	4		0	4 Errors 1	3 Errors 2	2 Errors 3	1 Error 4	0 Errors 5	
Total									Max.=30



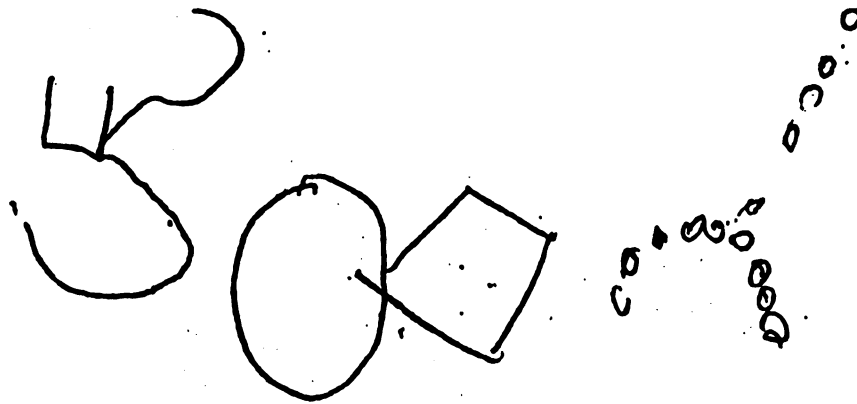
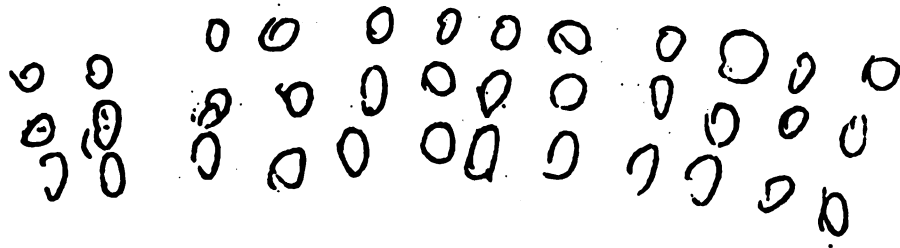
APPENDIX D

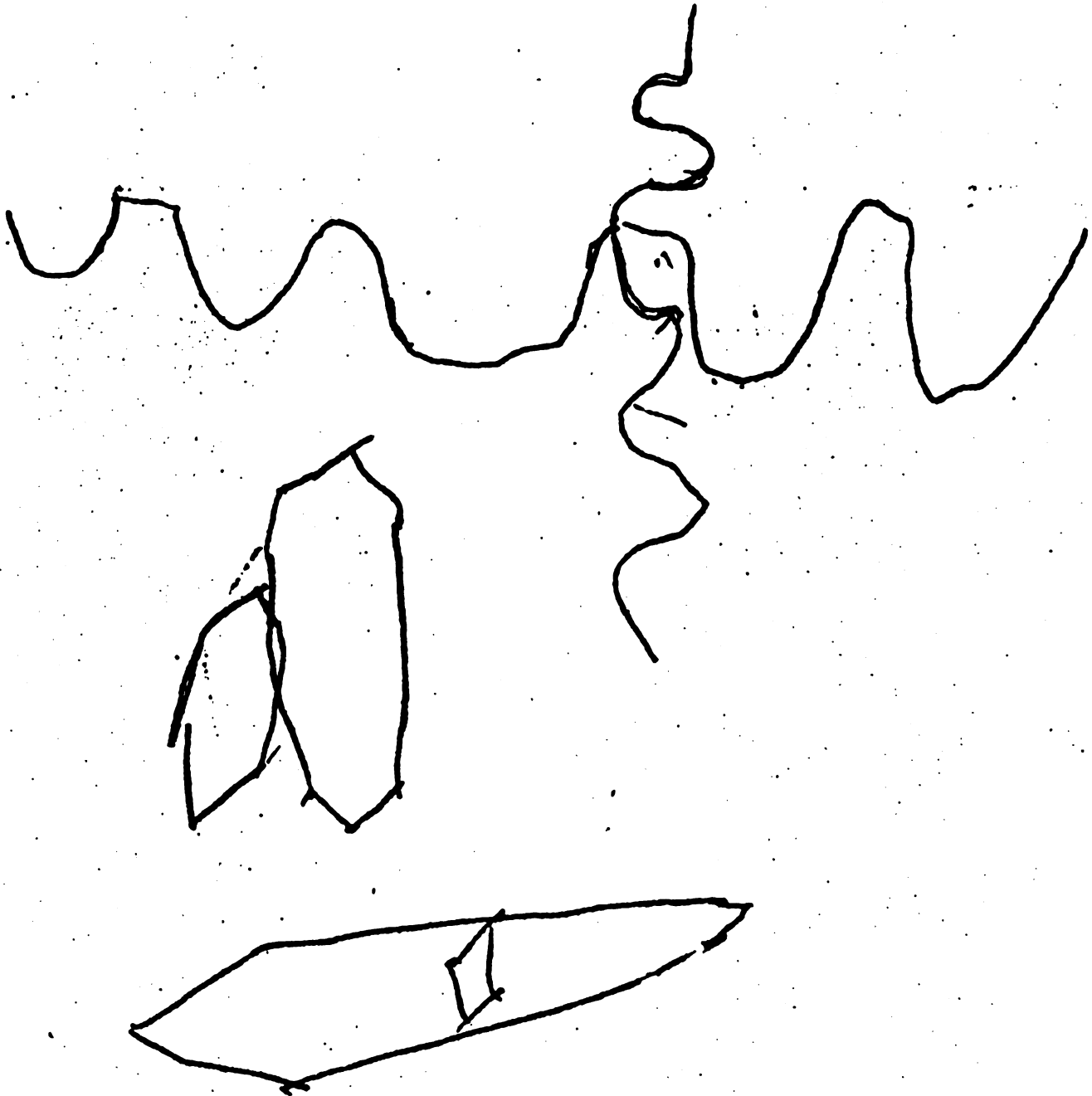
BENDER VISUAL MOTOR GESTALT TEST PROTOCOL



APPENDIX D

BENDER VISUAL MOTOR GESTALT TEST PROTOCOL





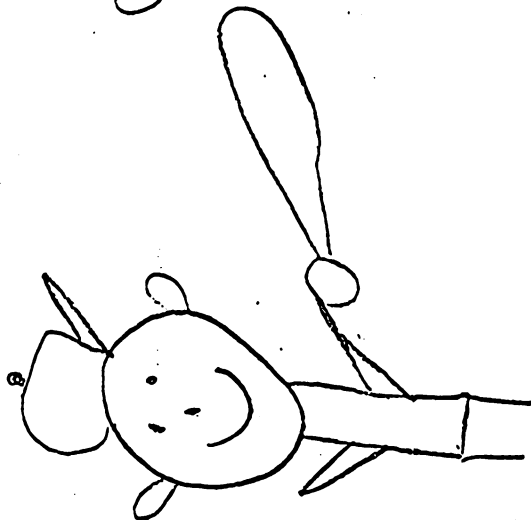
APPENDIX E

KINETIC FAMILY DRAWING

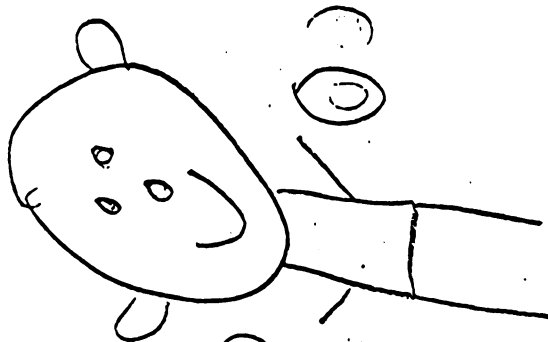


APPENDIX E

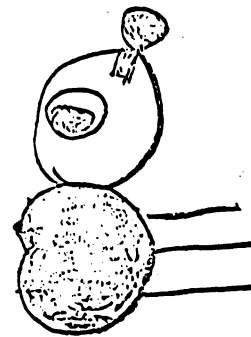
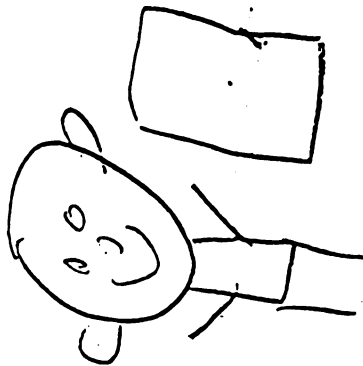
① we playing football



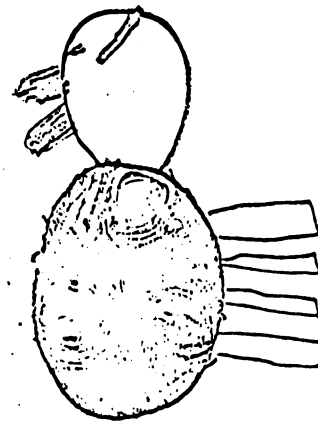
② My Mom doing the dishes



③ my dad give to the other



④ my dog



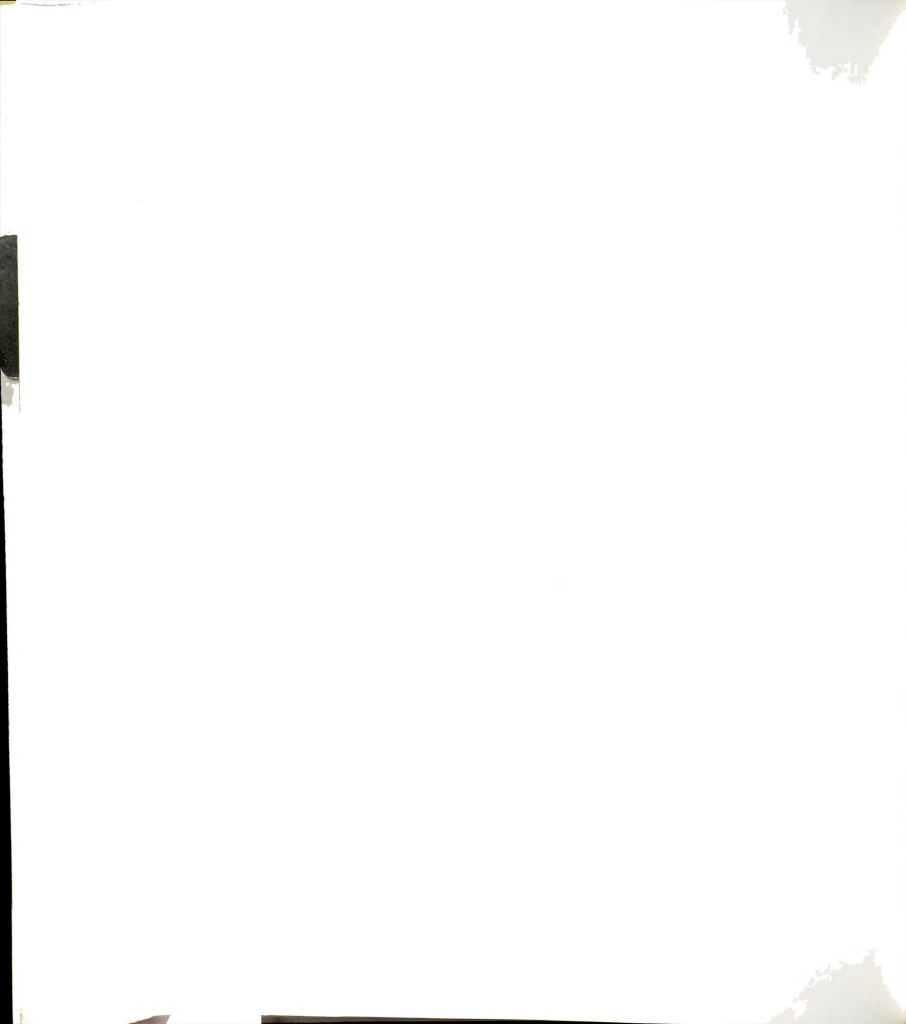
⑤ my pig

KINETIC FAMILY DRAWING



APPENDIX F

WIDE RANGE ACHIEVEMENT TEST PROTOCOL



APPENDIX F

WIDE RANGE ACHIEVEMENT TEST PROTOCOL

WIDE RANGE ACHIEVEMENT TEST

COPYRIGHT, 1965 by
Guidance Associates
1526 Gilpin Avenue
Wilmington, Delaware

Reading, Spelling, Arithmetic from Pre-School to College
By J. F. Jastak, S. W. Bijou, S. R. Jastak

Printed in U.S.A.
1937, 1946, 1963
Revised Edition
1965

Name SCOTT JONES Birthdate 8-8-68 M. F. Chron. Age 6-9
School _____ Grade 1 Reading Score 25 Grade 1.2 Stand. Sc. 92 %ile 30
Referred by _____ Spelling Score 20 Grade 1.2 Stand. Sc. 92 %ile 30
Date 3-26-75 Examiner [Signature] Arithmetic Score 19 Grade 1.6 Stand. Sc. 99 %ile 47

Percentiles and Standard Scores corresponding to grade ratings and age may be found in the Manual.

Level I—Spelling—Grade Norms.

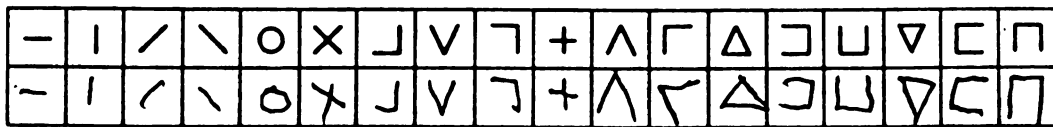
Score Grade	Score Grade	Score Grade	Score Grade	Score Grade
1 N.9	10 K.4	20 1.5	34 2.0	40 2.7
2 N.9	10 K.4	24 1.6	38 2.1	44 2.8
3 P.1.1	14 K.6	28 1.7	42 2.2	48 2.9
4 P.1.1	14 K.6	32 1.8	46 2.3	52 3.0
5 P.1.2	16 K.7	36 1.9	50 2.4	56 3.1
6 P.1.2	16 K.7	40 2.0	54 2.5	60 3.2
7 P.1.2	16 K.7	44 2.1	58 2.6	64 3.3
8 P.1.2	16 K.7	48 2.2	62 2.7	68 3.4
9 K.1	18 Gr.1.0	52 2.3	66 2.8	72 3.5
10 K.1	18 Gr.1.0	56 2.4	70 2.9	76 3.6
11 K.1	18 Gr.1.0	60 2.5	74 3.0	80 3.7

Level II—Spelling—Grade Norms.

Score Grade	Score Grade	Score Grade	Score Grade	Score Grade
0 K.6	11 4.0	21 4.7	31 5.0	41 5.4
1 K.6	11 4.0	25 4.8	35 5.1	45 5.5
2 Gr.1.0	13 4.0	29 4.9	39 5.2	49 5.6
3 1.2	14 4.0	33 5.0	43 5.3	53 5.7
4 1.2	14 4.0	37 5.1	47 5.4	57 5.8
5 1.2	14 4.0	41 5.2	51 5.5	61 5.9
6 1.2	14 4.0	45 5.3	55 5.6	65 6.0
7 1.2	14 4.0	49 5.4	59 5.7	69 6.1
8 1.2	14 4.0	53 5.5	63 5.8	73 6.2
9 1.2	14 4.0	57 5.6	67 5.9	77 6.3
10 1.2	14 4.0	61 5.7	71 6.0	81 6.4

Spelling Scores

Level I	Cumul. Score	Level II	Cumul. Score
Test	Score	Test	Score
Copying	1 point	Copying	4-9
1 point	1	10-17	2
per mark	18	18	3
Name	1 letter	Name	1 letter
1 letter	10	2 letters	4
2 letters	20	Spelling	1 point
Spelling	1 point	per word	6
1 point	21	per word	10
per word	65	per word	81



Name Scott 31 _____
1 g 16 _____ 32 _____
2 cx 17 _____ 33 _____
3 h 18 _____ 34 _____

LEVEL I

cat sea red to big work book eat was him how
then open letter jar deep even spell awake block size
weather should lip finger tray felt stalk cliff lame struck

Page 2. Arithmetic LEVEL I, Oral Part

3 pennies, spend 1? 2; 3 + 4 apples? 7; 9 marbles, lose 3? 6

Written part.

1 + 1 = 2 6 5 3 2
4 - 1 = 5 + 2 = 8 - 3 = 8 + 4 0
4 5 2
1 3 7
+ 2 4 5
6 + 2 = 8
Z H I Q S E B O
B O S E R T H P I U Z Q



APPENDIX G

SEVERITY OF PATHOLOGY SCALE



APPENDIX G

SEVERITY OF PATHOLOGY SCALE

Please estimate this child's functioning on the following continua.

Social Adjustment:

1	2	3	4	AVERAGE RANGE	5	6	7	8	9	10
anti-social										overly conforming

Inhibitory Functioning:

1	2	3	4	5	6	7	8	9	10
very impulsive									very inhibited

Behavioral Control:

1	2	3	4	5	6	7	8	9	10
hyper-active									over-controlled

Attentional Processes:

1	2	3	4	5	6	7	8	9	10
distractible									highly & narrowly focused

Perceptual-Motor Development

1	2	3	4	5	6	7	8	9	10
retarded									accelerated

Cognitive Organization:

1	2	3	4	5	6	7	8	9	10
Disorganized									over-organized, stereotypic

Relative Academic Functioning:

1	2	3	4	5	6	7	8	9	10
under-achieving									over-achieving

Overall Consistency of Functioning:

1	2	3	4	5	6	7	8	9	10
very spotty									very even

Overall Functioning:

1	2	3	4	5	6	7	8	9	10
mal-adjusted									superior adjustment



APPENDIX H

RATING SHEET:

DIAGNOSES

CONFIDENCE

SPECIAL EDUCATION TREATMENT RECOMMENDATION



APPENDIX H

RATING SHEET

Please indicate your choice of diagnosis using both classification systems.

MANDATORY SPECIAL EDUCATION SYSTEM

- ☐ not eligible for special education
- ☐ educable mentally impaired
- ☐ learning disabled
- ☐ emotionally impaired

CLASSICAL SYSTEM

- ☐ functioning within normal limits
- ☐ developmental personality deviation
- ☐ minimal brain dysfunction
- ☐ neurotic reaction
- ☐ behavior disorder
- ☐ other (specify) _____

Please rate the degree of confidence you feel in the diagnostic judgments you have made.

1	2	3	4	5	6	7	8	9	10
very									very
unsure									confident

Keeping in mind the diagnostic judgments you have made, try to choose the best of the following administrative educational programs for this particular child.

- ☐ enrollment in regular day class; no special education instruction, materials, or equipment.
- ☐ special education instructional materials and equipment only; enrollment in regular day class.
- ☐ special education instructional materials and equipment plus special education consultative services to regular teacher only; enrollment in regular day class.
- ☐ itinerant special education tutors; enrollment in regular day class.
- ☐ special education resource room and teacher; enrollment in regular day class.
- ☐ part-time special day class where enrolled; receive some academic instruction in regular day class.
- ☐ self-contained special day class where enrolled; receive no academic instruction in a regular day class.
- ☐ combination regular and special day school; receive no academic instruction in a regular day class.

APPENDIX I

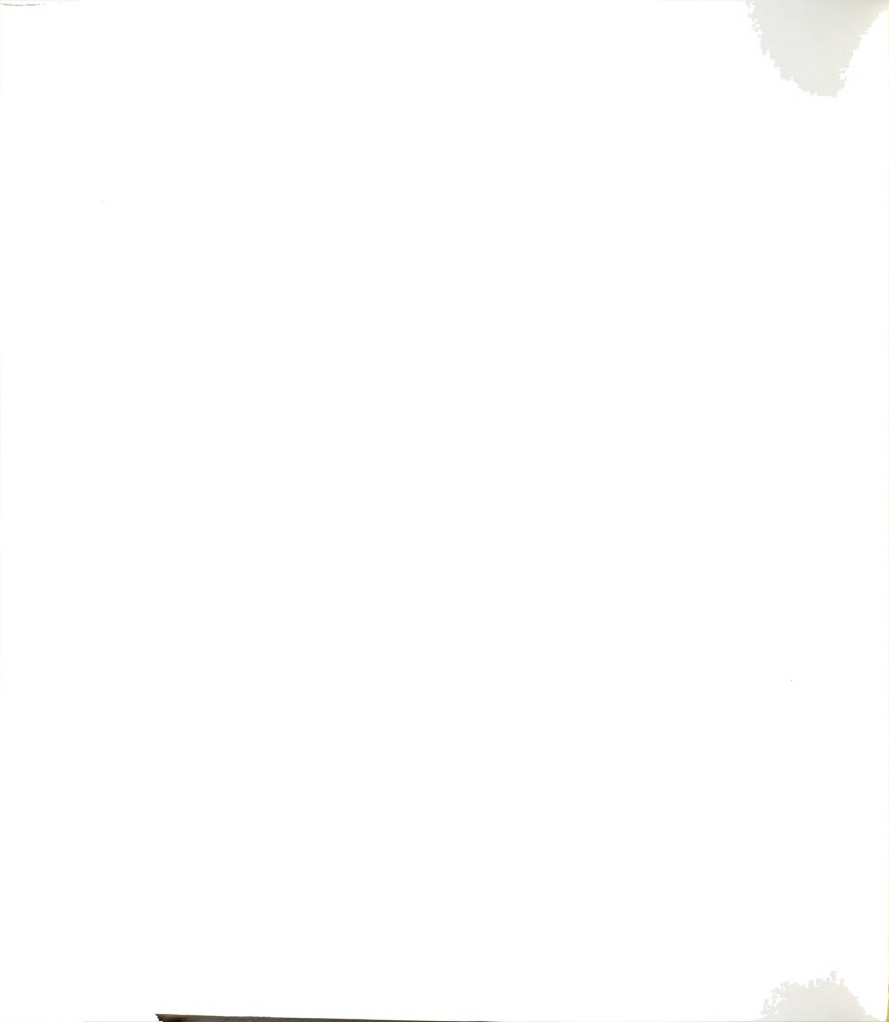
RATING SHEET:

REFERRAL FOR MEDICATION

THEORETICAL ORIENTATION

TRAINING

EXPERIENCE



APPENDIX I
RATING SHEET

Would you refer this child to his pediatrician for determination of medication possibilities?

_____ yes _____ no

Please indicate the degree to which you would agree or disagree with the following statements.

"If you have eliminated the behavior, you have usually solved the problem." (as opposed to the idea that underlying motivation will, very likely, cause symptom substitution if only surface behavior is treated)

1	2	3	4	5	6
strongly disagree					strongly agree

"The examination of early childhood experience is essential to the effective treatment of behavior/emotional problems in children." (as opposed to the idea that these problems can usually be treated effectively with a knowledge of current reinforcement patterns only)

1	2	3	4	5	6
strongly disagree					strongly agree

"Projective personality instruments are very useful in the study and treatment of childhood behavior/emotional problems." (as opposed to the idea that these instruments provide little information of practical use in treatment)

1	2	3	4	5	6
strongly disagree					strongly agree

"Most people need some kind of psychotherapeutic help at some time in their life." (as opposed to the idea that only a very small percentage of people need this kind of help)

1	2	3	4	5	6
strongly disagree					strongly agree

Please rate your theoretical preference along the following continuum.

1	2	3	4	5	6	7	8	9	10
strongly behavioristic									strongly psychodynamic

Please write the number of years you have been practicing in school psychology since your graduation from training: _____ years.

Was your graduate training primarily in _____ school psychology or in _____ clinical psychology?



APPENDIX J

RAW DATA FREQUENCY DISTRIBUTIONS

APPENDIX J

Table J-1
Raw Data Frequency Distributions

Length of Experience

Years	0-1	2-3	4-5	6-7	8-9	10-12	13-18
Frequency							
Negative	2	7	1	1	2	2	2
Positive	2	3	2	0	2	2	4
Behavioral	1	6	3	2	1	2	1
Total	5	16	6	3	5	6	7
Grand Rapids	2	1	5	0	2	2	0

Theoretical Orientation

Score	5-10	11-14	15-18	19-22	23-26	27-30
Frequency						
Negative	1	1	2	6	7	0
Positive	1	5	2	2	4	1
Behavioral	1	3	5	1	4	2
Total	3	9	9	9	15	3
Grand Rapids	0	1	2	3	5	1

Confidence

Rating	1-2	3	4	5	6	7	8	9-10
Frequency								
Negative	0	1	2	1	2	2	4	5
Positive	0	0	1	1	0	6	3	4
Behavioral	0	0	1	1	3	1	9	1
Total	0	1	4	3	5	9	16	10
Grand Rapids	0	3	1	1	1	2	2	2

Severity of Pathology

Score	4.5-7.5	8.5-11.5	12.5-15.5	16.5-19.5	20.5-24.0
Frequency					
Negative	3	6	6	2	0
Positive	3	6	3	3	0
Behavioral	1	2	8	3	2
Total	7	14	17	8	2
Grand Rapids	1	4	4	2	1

Special Education Treatment Recommendation

Score	1	2	3	4	5	6	7	8
Frequency								
Negative	7	1	7	0	0	1	1	0
Positive	2	3	3	3	2	1	1	0
Behavioral	5	1	2	3	5	0	0	0
Total	14	5	12	6	7	2	2	0
Grand Rapids	1	0	2	1	7	1	0	0

Table J-1 (Continued)

Classical Diagnosis

Diagnosis	Normal	Pathology
Frequency		
Negative	7	10
Positive	9	6
Behavioral	4	12
Total	20	28
Grand Rapids	2	10

Mandatory Diagnosis

Diagnosis	"Not Eligible"	Pathology
Frequency		
Negative	12	5
Positive	10	5
Behavioral	10	6
Total	32	16
Grand Rapids	4	8

Referral for Medication

Referral	Yes	No
Frequency		
Negative	7	10
Positive	5	10
Behavioral	8	8
Total	20	28
Grand Rapids	6	6

Graduate Training

Program	School	Clinical
Frequency		
Negative	15	2
Positive	8	6
Behavioral	13	3
Total	36	11
Grand Rapids	11	1

APPENDIX K

RANDOM GROUP TESTS OF ASSOCIATION

APPENDIX K

Table K-1

Random Group Chi Square Tests of Association with
all Treatment Conditions Combined (T_1, T_2 , and T_3 ; $n=48$)

"Classical" Diagnosis	Referral for Medication		row total
	yes	no	
normal	n=6	n=14	n=20
	row % 30.0	row % 70.0	% = 41.7
	col % 30.0	col % 50.0	
	tot % 12.5	tot % 29.2	
pathology	n=14	n=14	n=28
	row % 50.0	row % 50.0	% = 58.3
	col % 70.0	col % 50.0	
	tot % 29.2	tot % 29.2	
column total	n=20	n=28	
	% = 41.7	% = 58.3	p = .276

"Mandatory" Diagnosis	Referral for Medication		row total
	yes	no	
"not eligible"	n=9	n=23	n=32
	row % 28.1	row % 71.9	% = 66.7
	col % 45.0	col % 82.1	
	tot % 18.8	tot % 47.9	
pathology	n=11	n=5	n=16
	row % 68.8	row % 31.3	% = 33.3
	col % 55.0	col % 17.9	
	tot % 22.9	tot % 10.4	
column total	n=20	n=28	
	% = 41.7	% = 58.3	p = .017

"Mandatory" Diagnosis	"Classical" Diagnosis		row total
	normal	pathology	
"not eligible"	n=18	n=14	n=32
	row % 56.3	row % 43.8	% = 66.7
	col % 90.0	col % 50.0	
	tot % 37.5	tot % 29.2	
pathology	n=2	n=14	n=16
	row % 12.5	row % 87.5	% = 33.3
	col % 10.0	col % 50.0	
	tot % 4.2	tot % 29.2	
column total	n=20	n=28	
	% = 41.7	% = 58.3	p = .010

Table K-2

Random Group Fisher's Exact Tests Within the
Negative Treatment Condition (n=17)

"Classical" Diagnosis	Referral for Medication		row total
	yes	no	
normal	n=2	n=5	n=7
	row % 28.6	row % 71.4	% = 41.2
	col % 28.6	col % 50.0	
	tot % 11.8	tot % 29.4	
pathology	n=5	n=5	n=10
	row % 50.0	row % 50.0	% = 58.8
	col % 71.4	col % 50.0	
	tot % 29.4	tot % 29.4	
column total	n=7	n=10	
	% = 41.2	% = 58.8	p = .354

"Mandatory" Diagnosis	Referral for Medication		row total
	yes	no	
"not eligible"	n=4	n=8	n=12
	row % 33.3	row % 66.7	% = 70.5
	col % 57.1	col % 80.0	
	tot % 23.5	tot % 47.1	
pathology	n=3	n=2	n=5
	row % 60.0	row % 40.0	% = 29.4
	col % 42.9	col % 20.0	
	tot % 17.6	tot % 11.8	
column total	n=7	n=10	
	% = 41.2	% = 58.8	p = .314

"Mandatory" Diagnosis	"Classical" Diagnosis		row total
	normal	pathology	
"not eligible"	n=7	n=5	n=12
	row % 58.3	row % 41.7	% = 70.6
	col % 100	col % 50.0	
	tot % 41.2	tot % 29.4	
pathology	n=0	n=5	n=5
	0	row % 100	% = 29.4
	0	col % 50.0	
	0	tot % 29.4	
column total	n=7	n=10	
	% = 41.2	% = 58.8	p = .041

Table K-3

Random Group Fisher's Exact Tests Within the
Positive Treatment Condition (n=15)

"Classical" Diagnosis	Referral for Medication		row total
	yes	no	
normal	n=2	n=7	n=9
	row % 22.2	row % 77.8	% = 60.0
	col % 40.0	col % 70.0	
	tot % 13.3	tot % 46.7	
pathology	n=3	n=3	n=6
	row % 50.0	row % 50.0	% = 40.0
	col % 60.0	col % 30.0	
	tot % 20.0	tot % 20.0	
column total	n=5	n=10	
	% = 33.3	% = 66.7	p = .287

"Mandatory" Diagnosis	Referral for Medication		row total
	yes	no	
"not eligible"	n=2	n=8	n=10
	row % 20.0	row % 80.0	% = 66.7
	col % 40.0	col % 80.0	
	tot % 13.3	tot % 53.3	
pathology	n=3	n=2	n=5
	row % 60.0	row % 40.0	% = 33.3
	col % 60.0	col % 20.0	
	tot % 20.0	tot % 13.3	
column total	n=5	n=10	
	% = 33.3	% = 66.7	p = .167

"Mandatory" Diagnosis	"Classical" Diagnosis		row total
	normal	pathology	
"not eligible"	n=8	n=2	n=10
	row % 80.0	row % 20.0	% = 66.7
	col % 88.9	col % 33.3	
	tot % 53.3	tot % 13.3	
pathology	n=1	n=4	n=5
	row % 20.0	row % 80.0	% = 33.3
	col % 11.1	col % 66.7	
	tot % 6.7	tot % 26.7	
column total	n=9	n=6	
	% = 60.0	% = 40.0	p = .047

Table K-4
Random Group Fisher's Exact Tests Within the
Behaviorally Oriented Teacher Referral Condition (n=16)

"Classical" Diagnosis	Referral for Medication		row total
	yes	no	
normal	n=2	n=2	n=7
	row % 50.0	row % 50.0	% = 25.0
	col % 25.0	col % 25.0	
	tot % 12.5	tot % 12.5	
pathology	n=6	n=6	n=12
	row % 50.0	row % 50.0	% = 75.0
	col % 75.0	col % 75.0	
	tot % 37.5	tot % 37.5	
column total	n=8	n=8	
	% = 50.0	% = 50.0	p = .715

"Mandatory" Diagnosis	Referral for Medication		row total
	yes	no	
"not eligible"	n=3	n=7	n=10
	row % 30.0	row % 70.0	% = 62.5
	col % 37.5	col % 87.5	
	tot % 18.8	tot % 43.8	
pathology	n=5	n=1	n=6
	row % 83.3	row % 16.7	% = 37.5
	col % 62.5	col % 12.5	
	tot % 31.3	tot % 6.3	
column total	n=8	n=8	
	% = 50.0	% = 50.0	p = .059

"Mandatory" Diagnosis	"Classical" Diagnosis		row total
	normal	pathology	
"not eligible"	n=3	n=7	n=10
	row % 30.0	row % 70.0	% = 62.5
	col % 75.0	col % 58.3	
	tot % 18.8	tot % 43.8	
pathology	n=1	n=5	n=6
	row % 16.7	row % 83.3	% = 37.5
	col % 25.0	col % 41.7	
	tot % 6.3	tot % 31.3	
column total	n=4	n=12	
	% = 25.0	% = 75.0	p = .511

REFERENCES

REFERENCES

- Ash, P. The reliability of psychiatric diagnoses. Journal of Abnormal and Social Psychology, 1949, 44, 272-277.
- Auffrey, J., & Robertson, M. Case history information and examiner experience as determinants of scoring variability on Wechsler intelligence tests. Proceedings of the 80th Annual Convention of the American Psychological Association, 1972, 553-554.
- Babad, E. Y., Mann, M., & Mar-Hayim, M. Bias in scoring the WISC sub-tests. Journal of Consulting and Clinical Psychology, 1975, 43, 268.
- Baker, J. P., & Crist, J. L. Teacher expectancies: A review of the literature. In J. D. Elashoff & R. E. Snow (Eds.), Pygmalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971.
- Barber, T. X., & Silver, M. J. Fact, fiction and the experimenter bias effect. Psychological Bulletin, 1968, 70, 1-29.
- Beal, A. Biased therapists: The effects of prior exposure to case history material on the therapist's attitudes and behavior toward patients. Unpublished doctoral dissertation, Syracuse University, 1969.
- Beasley, D. S., & Manning, J. I. Experimenter bias and speech pathologists' evaluation of children's language skills. Journal of Communication Disorders, 1973, 6, 93-101.
- Beez, W. V. Influence of biased psychological reports on teacher behavior and pupil performance. Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, 3, 605-606.

- Bender, L. Instructions for the use of visual motor Gestalt test. New York: American Orthopsychiatric Association, 1946.
- Bloom, R., & Tesser, A. On reducing experimenter bias: The effects of forewarning. Canadian Journal of Behavioral Science, 1971, 3, 198-208.
- Boekel, N. The influence of teacher expectations on the performance of the educable mentally retarded. Focus on Exceptional Children, 1969, 1, 6-10.
- Brophy, J. E., & Good, T. L. Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. Journal of Educational Psychology, 1970, 61, 365-374.
- Brophy, J. E., & Good, T. L. Teacher-student relationships: Causes and consequences. New York: Holt, Rinehart, & Winston, 1974.
- Brown, W. The influence of student information on the formulation of teacher expectancy. Unpublished doctoral dissertation, Indiana University, 1969.
- Burns, R. C., & Kaufman, S. H. Actions, styles, and symbols in Kinetic family drawings (K-F-D): An interpretive manual. New York: Brunner/Mazel, 1972.
- Chaikin, A., Sigler, E., & Derlega, V. Non-verbal mediators of teacher expectancy effects. Unpublished manuscript, cited in J. E. Brophy & T. L. Good, Teacher-student relationships: Causes and consequences. New York: Holt, Rinehart, & Winston, 1974.
- Claiborn, W. L. Expectancy effects in the classroom: A failure to replicate. Journal of Educational Psychology, 1969, 60, 377-383.
- Conn, L., Edwards, C., Rosenthal, R., & Crowne, D. Perception of emotion and response to teacher's expectancy by elementary school children. Psychological Reports, 1968, 22, 27-34.
- Critchley, D. L. The relationship between the induced set of psychiatric diagnostic labels and close mindedness on the perception of child behavior among baccalaureate students in nursing. Unpublished doctoral dissertation, New York University, 1970.

- Dana, J. M., & Dana, R. H. Experimenter bias or task bias? Perceptual and Motor Skills, 1969, 29, 8.
- Dangel, H. L. Biasing effect of pretest referral information on WISC scores of mentally retarded children. American Journal of Mental Deficiency, 1972, 77, 354-359.
- Dunn, L. M. Exceptional children in the schools: Special education in transition (2nd edition). New York: Holt, Rinehart, & Winston, 1973.
- Dusek, J. B. Do teachers bias children's learning? Review of Educational Research, 1975, 45, 661-684.
- Education for all Handicapped Children Act. Public Law 94-142, November 29, 1975.
- Egeland, B. Examiner expectancy: Effects on the scoring of the WISC. Psychology in the Schools, 1969, 6, 313-316.
- Elashoff, J. D., & Snow, R. E. (Eds.). Pygmalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971.
- Fielder, W., Cohen, R., & Feeney, S. An attempt to replicate the teacher expectancy effect. Psychological Reports, 1971, 29, 1223-1228.
- Flanagan, J. C. Test of general ability: Technical report. Chicago: Science Research Associates, 1960.
- Fleming, E. S., & Anttonen, R. G. Teacher expectancy or my fair lady. American Educational Research Journal, 1971a, 8, 241-252.
- Fleming, E. S., & Anttonen, R. G. Teacher expectancy as related to the academic and personal growth of primary-age children. Monographs of the Society for Research in Child Development, 1971b, 36.
- Foster, G. G., Ysseldyke, J. E., & Reese, J. H. I wouldn't have seen it if I hadn't believed it. Exceptional Children, 1975, 41, 469-473.
- Gillingham, W. H. An investigation of examiner influence on Wechsler Intelligence Scale for Children scores. Unpublished doctoral dissertation, Michigan State University, 1970.

- Goldfarb, A. Reliability of diagnostic judgments by psychologists. Journal of Clinical Psychology, 1959, 15, 392-396.
- Grieger, R. M., & Richards, H. C. Prevalence and structure of behavior symptoms among children in special education and regular classroom settings. Journal of School Psychology, 1976, 14, 27-38.
- Grosz, H. J., & Grossman, K. G. The sources of observer variation and bias in clinical judgments: The item of psychiatric history. Journal of Nervous and Mental Disease, 1964, 138, 105-113.
- Grosz, H. J., & Grossman, K. G. Clinician's response style: A source of variation and bias in clinical judgments. Journal of Abnormal Psychology, 1968, 73, 207-214.
- Guskin, S. L. Dimensions of judged similarity among deviant types. American Journal of Mental Deficiency, 1963a, 68, 218-224.
- Guskin, S. L. Measuring the strength of the stereotype of the mental defective. American Journal of Mental Deficiency, 1963b, 67, 569-575.
- Herish, J. B. Effects of referral information on testers. Journal of Consulting and Clinical Psychology, 1971, 37, 116-122.
- Hipskind, N., & Rintelmann, W. Effects of experimenter bias upon pure tone and speech audiometry. Journal of Audiological Research, 1969, 9, 298-305.
- Hobbs, N. The futures of children. San Francisco: Jossey-Bass, 1975.
- Hollinger, C. S., & Jones, R. L. Community attitudes toward slow learners and mental retardates: What's in a name? Mental Retardation, 1970, 8, 19-23.
- Ingraham, L. H., & Harrington, G. M. Psychology of the scientist: SVI. Experience of E as a variable in reducing experimenter bias. Psychological Reports, 1966, 19, 455-461.



- Jacobs, J., & DeGraaf, C. Expectancy and race: Their influences upon the scoring of individual intelligence tests. Paper presented at the Annual Meeting of the American Educational Research Association, 1973.
- Jastak, J. F., Bijou, S. W., & Jastak, S. R. Wide range achievement test: Reading, spelling, arithmetic from pre-school to college. Wilmington, Delaware: Guidance Associates, 1965.
- Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.
- Jones, R. L. Labels and stigma in special education. Exceptional Children, 1972, 38, 553-564.
- Jose, J., & Cody, J. J. Teacher-pupil interaction as it relates to attempted changes in teacher expectancy of academic ability and achievement. American Educational Research Journal, 1971, 8, 39-49.
- Katz, M. M., Cole, J. O., & Lowery, H. A. Studies of the diagnostic process: The influence of symptom perception, past experience and ethnic background on diagnostic decisions. American Journal of Psychiatry, 1969, 125, 937-947.
- Kent, R. N., O'Leary, K. D., Diament, C., & Dietz, A. Expectation biases in observational evaluation of therapeutic change. Journal of Consulting and Clinical Psychology, 1974, 42, 774-780.
- Kessel, N., & Shepherd, M. Neurosis in hospital and general practice. Journal of Mental Science, 1962, 108, 159-166.
- Kester, S., & Letchworth, G. Communication of teacher expectations and their effects on achievement and attitudes of secondary school students. The Journal of Educational Research, 1972, 66, 51-55.
- Langer, E. J., & Abelson, R. P. A patient by any other name . . .: Clinician group difference in labeling bias. Journal of Consulting and Clinical Psychology, 1974, 42, 4-9.

- Larrabee, L., & Kleinsasser, L. The effect of experimenter bias on WISC performance. Unpublished manuscript, St. Louis: Psychological Associates, 1967.
- Lasky, D. I., Felice, A., Moyer, R. C., Buddington, J. F., & Elliot, E. S. Examiner effects with the Peabody Picture Vocabulary Test. Journal of Clinical Psychology, 1973, 29, 456-457.
- Lee, S. D., & Temerlin, M. K. Social class, diagnosis, and prognosis for psychotherapy. Psychotherapy: Theory, Research and Practice, 1970, 7, 181-185.
- MacFarlane, J., Allen, L., & Honzik, M. A developmental study of the behavior problems of normal children between twenty-one months and fourteen years. Berkeley: University of California Press, 1954.
- Mandatory Special Education Act. Public Acts of Michigan, 1971, No. 198.
- Marwit, S. J. An investigation of the communication of tester bias by means of modeling. Unpublished doctoral dissertation, State University of New York at Buffalo, 1968.
- Marwit, S. J., & Marcia, J. E. Tester bias and response to projective instruments. Journal of Consulting Psychology, 1967, 31, 253-258.
- Masling, J. M. The effects of warm and cold interaction on the administration and scoring of an intelligence test. Journal of Consulting Psychology, 1959, 23, 336-341.
- Masling, J. M. Differential indoctrination of examiners and Rorschach responses. Journal of Consulting Psychology, 1965, 29, 198-201.
- Medinnus, G., & Unruh, R. Teacher expectations and verbal communication. Paper presented at the Annual Meeting of the Western Psychological Association, 1971.
- Mehlman, B. The reliability of psychiatric diagnoses. Journal of Abnormal and Social Psychology, 1952, 47, 557-578.
- Meichenbaum, D., Bowers, K., & Ross, R. A behavioral analysis of teacher expectancy effect. Journal of Personality and Social Psychology, 1969, 13, 306-316.

- Meitus, I. J., Ringel, R. L., House, A. S., & Hotchkiss, J. C. Clinical bias in evaluating speech proficiency. British Journal of Disorders of Communication, 1973, 8, 146-151.
- Meyers, C. E., Sitkei, E. G., & Watts, C. A. Attitudes toward special education and the handicapped in two community groups. American Journal of Mental Deficiency, 1966, 71, 78-84.
- Morrow, H. W., Powell, G. D., & Ely, D. D. Placement or placebo: Does additional information change special education placement decisions? Journal of School Psychology, 1976, 14, 186-191.
- Pflugrath, J. Examiner influence in a group testing situation with particular reference to examiner bias. Unpublished master's thesis, University of North Dakota, 1962.
- Raffetto, A. M. Experimenter effects on subjects' reported hallucinatory experience under visual and auditory deprivation. Unpublished master's thesis, San Francisco State College, 1967.
- Rist, R. Student social class and teacher expectations: The self-fulfilling prophesy in ghetto education. Harvard Educational Review, 1970, 40, 411-451.
- Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.
- Rosenthal, R. Experimenter effects in behavioral research: Enlarged edition. New York: Halsted Press, 1976.
- Rosenthal, R. Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R. L. Rosnow (Eds.), Artifact in behavioral research. New York: Academic Press, 1969a, 181-277.
- Rosenthal, R. Task variations in studies of experimenter expectancy effects. Perceptual and Motor Skills, 1969b, 29, 9-10.
- Rosenthal, R., & Fode, K. L. The problem of experimenter outcome bias. In D. P. Ray (Ed.), Series research in social psychology, Symposia Studies Series, No. 8. Washington, D.C.: National Institute of Social and Behavioral Science, 1961.

- Rosenthal, R., & Fode, K. L. Psychology of the scientist: V. Three experiments in experimenter bias. Psychological Reports, 1963, 12, 491-511.
- Rosenthal, R., & Jacobson, L. Teachers' expectancies: Determinants of pupils' IQ gains. Psychological Reports, 1966, 19, 115-118.
- Rosenthal, R., & Jacobson, L. Pygmalion in the classroom: Teacher expectations and pupils' intellectual development. New York: Holt, Rinehart, & Winston, 1968.
- Rosenthal, R., & Rosnow, R. L. (Eds.). Artifact in behavioral research. New York: Academic Press, 1969.
- Rosenthal, R., & Rubin, D. B. Pygmalion reaffirmed. In J. D. Elashoff & R. E. Snow (Eds.), Pygmalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971.
- Rothbart, M., Dalfen, S., & Barrett, R. Effects of teacher's expectancy on student-teacher interaction. Journal of Educational Psychology, 1971, 62, 49-54.
- Rubovits, P. C., & Maehr, M. L. Pygmalion analyzed: Toward an explanation of the Rosenthal-Jacobson findings. Journal of Personality and Social Psychology, 1971, 19, 197-203.
- Sarbin, T. R. On the futility of the proposition that some people be labeled "mentally ill." Journal of Consulting Psychology, 1967, 31, 447-453.
- Sattler, J. M. Assessment of children's intelligence: Revised reprint. Philadelphia: W. B. Saunders, 1974.
- Sattler, J. M., Hillix, W. A., & Neher, L. A. Halo effect in examiner scoring of intelligence test responses. Journal of Consulting and Clinical Psychology, 1970, 34, 172-176.
- Sattler, J. M., & Theye, F. Procedural, situational, and interpersonal variables in individual intelligence testing. Psychological Bulletin, 1967, 68, 347-360.

- Sattler, J. M., & Winget, B. M. Intelligence testing procedures as affected by expectancy and IQ. Journal of Clinical Psychology, 1970, 26, 446-448.
- Saunders, B. T., & Vitro, F. T. Examiner expectancy and bias as a function of the referral process in cognitive assessment. Psychology in the Schools, 1971, 8, 168-171.
- Schroeder, H. E., & Kleinsasser, L. D. Examiner bias: A determinant of children's verbal behavior on the WISC. Journal of Consulting and Clinical Psychology, 1972, 39, 451-454.
- Shames, M. L., & Adair, J. G. Experimenter-bias as a function of the type and structure of the task. Paper presented at the Meeting of the Canadian Psychological Association, Ottawa, May, 1967.
- Shotel, J. R., Iano, R. P., & McGettigan, J. F. Teacher attitudes associated with the integration of handicapped children. Exceptional Children, 1972, 38, 677-683.
- Simon, W. E. Expectancy effects in the scoring of vocabulary items: A study of scorer bias. Journal of Educational Measurement, 1969, 6, 159-164.
- Snow, R. L. Unfinished Pygmalion. Contemporary Psychology, 1969, 14, 197-199.
- Strauss, M. E. Examiner expectancy: Effects on Rorschach experience balance. Journal of Consulting and Clinical Psychology, 1968, 32, 125-129.
- Stuart, R. Trick or treatment--How and when psychotherapy fails. Urbana, Ill.: Research Press, 1970.
- Szasz, T. The myth of mental illness. New York: Hoeber-Harper, 1961.
- Taylor, J. A. A personality scale of manifest anxiety. Journal of Abnormal and Social Psychology, 1953, 48, 285-290.
- Temerlin, M. K. Suggestion effects in psychiatric diagnosis. Journal of Nervous and Mental Disease, 1968, 147, 349-359.

- Temerlin, M. K., & Trousdale, W. W. The social psychology of clinical diagnosis. Psychotherapy: Theory, Research and Practice, 1969, 6, 24-29.
- Thorndike, R. L. Review of Robert Rosenthal and Lenore Jacobson, Pygmalion in the classroom. American Educational Research Journal, 1968, 5, 708-711.
- Tolor, A., & Brannigan, G. G. How different are students of school psychology and clinical psychology? Psychology in the Schools, 1976, 13, 279-283.
- Towbin, A. When are cookbooks useful? American Psychologist, 1960, 15, 119-123.
- 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 University, 1962.
- Wechsler, D. Manual for the Wechsler Intelligence Scale for Children--Revised. New York: Psychological Corporation, 1974.
- Wechsler Intelligence Scale for Children--Revised: Record form. New York: Psychological Corporation, 1974.
- Werry, J. S., & Quay, H. C. The prevalence of behavior symptoms in younger elementary school children. American Journal of Orthopsychiatry, 1971, 41, 136-143.



MICHIGAN STATE UNIV. LIBRARIES



31293000241178