LIPINSKI, JUDITH MACKAY. The Effects of Expectancy and Parenthood on Observations by Parents. (1974) Directed by: W. Larry Osborne. Pp. 82.

With the move from clinical settings to the "natural" environment, parents have become mediators of behavioral change with their children. In this role, parents often serve as both experimenters and observers. The present study investigated some of the parameters of parental observations such as sex of parent, reliability, effects of expectancy, and effects of parenthood.

Four sets of parents were trained on a six-category behavioral code via video tape. Parents then observed four, equivalent (equal number of positive and negative behaviors) 10-minute video tapes, each tape being a different child and a different condition. The four conditions used were: positive expectancy, negative expectancy, neutral, and own-child.

Given the small amount of training in observations (one hour), the parents proved to be reliable observers -- mean reliability of .79. There were no significant effects on the observational data due to sex of parent nor expectancy. In addition, the data were not significantly affected by parenthood -- i.e., observing one's own child versus observing another child (neutral tape). However, there were significant effects ($p < .05$) due to expectancy on global ratings of the children. The negative expectancy tapes received significantly more negative ratings than the

other tapes. The results were discussed in terms of the implications for utilization of parents as observers and the implications for working with parents of behavior-problem children.

THE EFFECTS OF EXPECTANCY AND PARENTHOOD ON OBSERVATIONS BY PARENTS

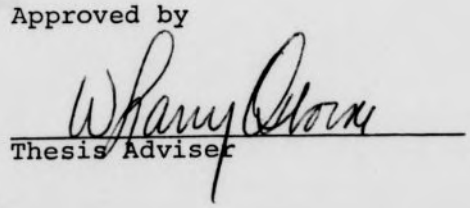
by

Judith MacKay Lipinski

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts in Education

Greensboro
1974

Approved by



Thesis Adviser

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of the Graduate School at The University of North Carolina at Greensboro.

Thesis Adviser W. Larry Gross

Committee Members John Christian Bleich

P. Scott Lawrence

July 3, 1974
Date of Acceptance by Committee

ACKNOWLEDGEMENTS

I would like to thank all of the members of my thesis committee -- Drs. Chris Busch, P. Scott Lawrence, and W. Larry Osborne -- for all their help throughout this research. In addition, I would like to thank my husband, David, who assisted with the video-taping, observing, and training of the parent-observers.

Further, I would like to thank all of the staff members of Henry Wiseman Kendall Center who facilitated this research -- especially Mr. Art Peoples and Ms. Ava Williams. A special thanks to Ms. Charma Pickett, who allowed us to use her treatment area and who prompted behaviors in the children.

And last, but certainly not least, I would like to thank all of the parents who served as observers for this study.

TABLE OF CONTENTS

Page

APPENDICES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Background

In recent years, behavior modifiers have turned their attention from the "artificial" settings of clinics and hospitals to the "natural" environments of classrooms and homes. That is, rather than conduct individual therapy with a child, therapists train mediators -- parents, teachers, peers, siblings -- to alter contingencies within the natural environment which in turn will change the child's behavior (Baer & Wolf, 1967; Patterson, McNeal, Hawkins, & Phelps, 1967; Ray, 1969; Stuart, 1971; Tharp & Wetzel, 1969; Wahler, 1970). In many such programs, parents (and teachers) act as both observers and experimenters (i.e., Hall, Axelrod, Tyler, Grief, Jones, & Robertson, 1972; Hall, Cristler, Cranston, & Tucker, 1970). However, the bias effects often attributed to experimenters (Rosenthal & Fode, 1963a, 1963b) and to observers (O'Leary, 1973) have not been systematically investigated with regard to parents.

The implication of some research (i.e., Johnson & Lobitz, 1972) is that parents can manipulate their children to behave in a favorable or unfavorable manner. A parent could make a child appear deviant during baseline

to justify treatment and then make the child appear improved to justify termination of treatment and to please the therapist. While this may be "intentional" error (Rosenthal, 1966), it would be difficult for a therapist to discriminate between this intentional error and unintentional error on the part of a parent. Despite the controversy over the existence of experimenter bias (Barber & Silver, 1968), it appears that a parent as experimenter can manipulate the behavior of the child in a fashion supportive of the hypothesized change in behavior from baseline to treatment.

The observer-experimenter parent being fully aware of the target behavior, the treatment and the expected outcome is susceptible to both experimenter and observer bias. The parent may possess certain expectancies about his child. In addition, a parent may be susceptible to what O'Leary (1973) terms observer drift -- that is, his definitions of the behaviors to be recorded may change over time from the original definitions. Furthermore, the parent often knows that no reliability measure is being taken; this in itself can reduce reliability (Reid, 1970). Finally, the parent receives feedback on his data from the therapist; this feedback tends to "shape" observations in the direction of the expected outcome (O'Leary, 1973).

Given the above strong possibilities for bias in observations by parents and the known inaccuracy of parents of retarded children to judge their child's development (Ewert & Green, 1957; Gorelick & Sandhu, 1967; Capobianco & Knox, 1964; Heriot & Schmickel, 1967), and the inaccuracy of some parents to correctly discriminate responses of their children (Dammann, 1973), more research should be conducted to determine the reliability of parental observations and the susceptibility of parents to experimenter and/or observational bias.

Statement of the Problem

Given the above rationale for investigating parents as experimenters and observers, the present research attempted to answer seven questions with regard to parents as observers:

1. Are parents reliable observers?
2. What is the effect of sex of parent on the reliability of parental observations?
3. What is the effect of own-child versus other-child status on the reliability of parental observers?
4. Are parents susceptible to observer bias by labelling?
5. What is the effect of sex of parent on observer bias by labelling?
6. Does own-child status bias parental observations as compared to other-child status?

7. What effect does sex of parent have on parental observations of their own child versus observations of other children?

To investigate questions 1, 2, and 3, two independent variables were utilized: sex of parent and status of child (own-child versus other-child). The dependent variable was reliability (agreement) between parent-observers and independent observers. Questions 4 and 5 were studied with two independent variables and two dependent variables; the independent variables were sex of parent and bias condition (neutral, positive, and negative) and the two dependent variables were number of positive behaviors recorded by the parent observers and the number of negative behaviors recorded by the parent observers. These same two dependent variables were used to research questions 6 and 7. The independent variables for these questions were sex of parent and status of child (own-child versus other-child).

Hypotheses

There were nine hypotheses to be tested:

1. Parents are as reliable as independent observers.

2. Fathers are more reliable observers than mothers.

While Capobianco and Knox (1964) found a significant difference between actual IQ scores and parental estimates of intelligence of their children, when they analyzed the

data by sex, they found mothers were quite inaccurate but that there was no difference between the father's estimates and the actual IQ scores of their children.

3. Parents are more reliable when observing other children than they are when observing their own children.

The assumption for this hypothesis is based on Dammann's (1973) finding of a difference between parental assessment of their own child's behavior and another child's behavior.

4. Parents record more positive behavior in a positive expectancy condition than in a neutral condition.

5. Parents record less negative behavior in a positive expectancy condition than in a neutral condition.

6. Parents record more negative behavior in a negative expectancy condition than in a neutral condition.

7. Parents record less positive behavior in a negative expectancy condition than in a neutral condition.

Hypotheses 4 through 7 are based on the assumption that parents are susceptible to observer bias just as trained observers are (Kass & O'Leary, 1970).

8. There is a difference in the number of positive behaviors recorded in the own-child condition versus the other-child condition.

9. There is a difference in the number of negative behavior recorded in the own-child condition versus the other-child condition.

Hypotheses 8 and 9 are based on the assumption that parents possess expectancies about their children which bias their observations of their own children as compared to other children.

Relevance of Findings

The determination of the reliability of parent observers is important to the field of clinical psychology, especially areas which use parents as observers in programs designed to modify child behavior. The move from the clinical to the natural setting was made without assessing the reliability of social agents as either experimenters or observers. Parents appear to be reliable observers, thus strengthening this move into the "natural" environment. Further, parents appear to be unbiased observers -- that is, they do not let expectations affect their data. However, more research is necessary to further investigate the reliability of parents in the "natural" setting with covert reliability measures.

Limitations of Findings

This research was conducted with parents of developmentally disabled (retarded) children. Differences may exist between these parents and parents of normal children which would limit the generalizability of the findings.

In addition, parents in this study were using a specific behavioral code and were observing video tapes. This may limit the generalizability of the findings to parental observations with other behavioral codes and parental observations of on-going behavior in the home. Further, the parents were aware that their accuracy was being assessed. This condition would not exist in the "natural" environment. Reliability tends to be lower when observers are unaware that their accuracy is being assessed (Lipinski & Nelson, 1974b; Reid, 1970). Therefore, parents may be less reliable when making observations at home.

CHAPTER II

REVIEW OF RELATED RESEARCH

Since the problem discussed in the Introduction has not been directly investigated, related research in the areas of experimenter and observer bias will be reviewed. First, experimenter bias -- its parameters and research data -- will be discussed. Then the tangential areas of teacher expectancy and labelling will be reviewed. The third major area to be presented is observer bias. Research relating to parental estimates of children's ability will be the fourth topic considered. Finally, parents as observers and experimenters will be explored.

Experimenter Bias Effects

Rosenthal (1966) aroused the current controversy concerning the experimenter bias effect. His basic proposition was that knowledge of expected results biases an experimenter's data in the direction of the expected results. Rosenthal reviewed numerous studies investigating the effect of particular experimenter attributes on subjects' responses. He outlined four basic factors or attributes that have been demonstrated to affect subjects' responses. The first area was biosocial factors such as sex, race, age and religion of the experimenter. While data show that these factors influence subjects' responses, it has not

yet been determined if the influence is a result of subjects' perceptions of the experimenter and behavior based on these perceptions or because experimenters with different biosocial attributes behave differently toward subjects. His second set of experimenter attributes affecting subjects' responses was psychosocial factors such as anxiety, need for approval, birth order, hostility, authoritarianism, intelligence and dominance. Social psychological attributes such as experimenter status and warmth were a third set of factors presented by Rosenthal. The final set of factors related to differential subject responses was situational variables -- for example, experimenter's prior contact with a subject, the amount and kind of prior experience of the experimenter and the physical setting. According to Rosenthal, any of the above factors alone or in interaction with the others may affect subjects' responses. Above and beyond these effects on subjects' responses, Rosenthal discussed other effects that may influence experimental data. The first effect he termed intentional error -- i.e., cheating or data fabrication. Rosenthal noted this intentional error to be especially prevalent among undergraduate psychology students who possibly feel that their grade is dependent upon turning in a "better report" -- one with expected data. Another prevalent "cheater" population according to Rosenthal was that

of the part-time survey-research interviewers. The second effect Rosenthal labelled unintentional error. He defined this term as error resulting from "highly effective and influential unprogrammed and unintended communication" (p. 402) which he hypothesized to be subtle, non-verbal cues. This unintentional error is seen by Rosenthal as the process by which experimenter biasing of data occurs. However, it is often difficult to determine whether the error, which results in biased data, is intentional or unintentional.

Rosenthal used two basic paradigms to investigate the experimenter bias effect. The first paradigm involved non-human subjects; for example, Rosenthal and Fode (1963a) randomly assigned naive rats to two groups of experimenters (undergraduate students). One group of student-experimenters was told that their rats were bred to be "maze-bright" animals; the other group of experimenters was informed their rats were "maze-dull." Both groups of experimenters then proceeded to test their rats on a maze-learning task. Experimenters told that their rats were bred for "maze-brightness" reported faster learning times for their animals than experimenters informed they were testing "maze-dull" rats.

The second paradigm involved human subjects. Typical of this paradigm was an experiment in which two groups of

undergraduate experimental psychology students served as experimenters (Rosenthal & Fode, 1963b). The student-experimenters were matched for grades in experimental psychology, assuming that the grades reflected the degree of sophistication of each student as an experimenter. Student-experimenters were trained to run human subjects (introductory psychology students) on a photo-rating task, often called person-perception task. The task required the subject to rate a neutral photograph on a scale from -5 to +5 in regard to a particular variable -- i.e., success or failure of the person. All experimenters were informed that the study was an attempt to replicate findings from previous research. One-half of the experimenters were told to expect high ratings (around +5) and the other half were told to expect low ratings. All experimenters were paid $1.00 per hour and were told that if they did a "good" job (that is, brought in data supportive of the hypothesis) they would receive $2.00 per hour. The authors reported a significant biasing effect ($p = .007$) -- both groups of experimenters reported data supportive of the expectancies they received.

One might hypothesize that motivation could account for the above results. Rosenthal (1963) himself admitted that motivation is an important variable. He stated that if the experimenter thinks he is being bribed to bias

(paid for certain results) or if the experimenter is extremely aware of his motivation to obtain certain data, then there is a tendency toward a significant "reverse" bias effect. That is, the experimenter brings in data in opposition to his expectancy; for example, if biased to obtain "high" ratings on the person-perception task, he would obtain ratings lower than an experimenter with no bias (control). Rosenthal also reported that under paid conditions with explicit bias directions, 4% of the data was accounted for by the bias effect; yet, under non-paid conditions with an expectancy statement, the bias effect accounted for 18% of the variance. However, he discussed one study in which the expectancy proved more powerful than motivation. Introductory psychology students were required to complete one experiment before they could continue to the other experiments required for the course. The first experiment was a maze-learning task; while the rats had to learn the maze before the student-experimenters could go any further, those student-experimenters given a "dull" expectancy obtained "dull" performance from their rats.

Numerous attempts have been made to replicate Rosenthal's findings of an experimenter bias effect. Miller (1972) and Page (1972) have both used Rosenthal's human-subject paradigm and his person-perception task.

In three studies, Page used psychiatric patients and personnel as experimenters and subjects. In the first study, nine male psychiatric patients served as experimenters and 63 male and female psychiatric patients were the subjects. Rosenthal's person-perception task was used with an enlarged rating scale from -10 to +10. Half of the patient-experimenters were given an expectancy of "successful" (high) ratings; the other half received no expectancy. Page reported no bias effect -- i.e., no significant differences between the two groups of experimenters. In the second study, three ward nurses acted as experimenters for 11 schizophrenic patients, both male and female. The same design was utilized but Page noted a reverse bias effect ($p = .06$); in other words, experimenters given a high expectancy reported significantly lower ratings than the experimenter given no expectancy. Ten male alcoholic patients were the experimenters in the third study with ten other alcoholic patients being the subjects. In this last study, a significant ($p < .01$) biasing effect appeared. However, the latter two studies which yielded significant effects involved smaller numbers of experimenters and subjects. The first study with a larger sample size resulted in no significant effects.

Miller (1972) used students as experimenters and subjects to investigate the effect of experimenter status and

competence on the experimenter bias effect. He trained eight law and business graduate students in influence techniques and in how to administer the person-perception task; in addition, he gave them specific training in how to behave as a competent experimenter and how to behave as an incompetent experimenter. Three competence conditions were utilized: 1) the student-experimenters were instructed to behave neither deliberately competently nor incompetently; 2) the student-experimenters were instructed to behave deliberately competently; and 3) the student-experimenters were instructed to behave deliberately incompetently. The high or low-status conditions were manipulated by signs on the doors to the experimental rooms; student-experimenters were unaware of this manipulation. All the student-experimenters participated in all six treatment conditions and all received expectancies prior to the start of the experiment. Miller reported that when the student-experimenters were acting neither competently nor incompetently, higher-status experimenters were slightly better at biasing the data (i.e., influencing the subjects) than lower-status experimenters. When behaving deliberately competently, the higher-status experimenters were significantly more successful at biasing; but when behaving deliberately incompetently, the lower-status experimenters were slightly more successful at biasing the data. The

author reported a significant interaction effect (status x competence); however, he accepts a probability level of .09 for significance. Since this research was a precisely designed study (rather than exploratory research), the author has increased the chance of a type I error -- that is, assuming there is an interaction effect when none may exist.

In addition to the above findings, Miller discovered a significant positive correlation ($r = .77$, $p < .002$) between the amount of the bias effect and the ambiguity of the task. Prior to the above experiment, he obtained "unbiased" or no-expectancy ratings for all the pictures. He used the standard deviation of the ratings for each picture as a measure of the ambiguousness of the picture. This ambiguity measure for each picture was then correlated with the amount of bias obtained for that picture and the correlation proved significant. In other words, the student-experimenters were more successful at influencing subjects (biasing the data) on pictures which were more ambiguous (more subjective).

Brager (1969) reported a similar finding. He assigned randomly-chosen, fourth-grade teachers from six schools to three treatment groups (two schools per group). The first group was a positive (PB) group; the second, a negative bias (NB) group and the third group was a neutral bias (OB)

or control group. All teachers were told they would be validating judgments rendered by a panel of other teachers. The first experimental task the teachers performed was to read a short story to their class; then, the students rated ten statements concerning the story -- statements regarding the characterizations, morals, etc. -- on a disagree-agree scale. The PB group expected to obtain "agree" ratings; the NB teachers expected to obtain "disagree" ratings; and the control teachers expected either "undecided" responses or random ratings. In the second task, the teachers showed their classes reproductions of five paintings; the students rated each painting on a dislike-like scale. Here, the PB group expected "like" ratings while the NB group expected "dislike" ratings and the control group had no specific expectation. In both of these tasks, there was a significant difference between ratings obtained by the three groups of teachers ($p < .01$) and this difference was in the direction of the bias in both cases. These two tasks could be described as rather subjective tasks for the students. The third task which the teachers used was the most objective task of the three and one which required little teacher-student interaction; it was a modified programmed lesson in math. PB teachers expected the lesson in math. PB teachers expected the lesson would prove to be an "effective learning device" but the NB teachers were

given the expectancy of an "ineffective learning device." No significant differences were found among the means of the three groups although the means were ordered from NB to OB to PB, with PB being the highest. Brager's results lend support to Miller's (1972) conclusion that the amount of the biasing effect is directly related to the ambiguity of the task.

Dusek (1971, 1972) has used a motor task with young children to investigate the experimenter bias effect. In his first study, Dusek employed 18 male college students as experimenters for 126 first- and second-grade subjects. The task involved taking marbles from an open bin and dropping them into a covered bin; the cover on the second bin had small holes so that the marbles needed to drop through one of the holes to be considered "in" the bin. The dependent measure was the rate (marbles/minute) of marbles dropped into the second bin. Three bias conditions were utilized: boys will perform better, girls will perform better, and no expectancy. With these male student experimenters, the only significant bias effect occurred with the "girls will perform better"-condition. The same motor task was used in Dusek's second study. Six female college students served as experimenters for 48 nine-year-old subjects. The subjects were evenly divided into low- and high-test anxious groups and evenly divided by sex.

One-half of the student-experimenters were told that girls would perform better than boys; the other half of the student-experimenters were given the opposite expectancy. The only significant biasing effect occurred when the low-anxious subjects were tested by student-experimenters biased toward girls.

Barber, Forgione, Chaves, Calverly, McPeake and Bowen (1969) presented five identical studies conducted at five different colleges -- all attempts to replicate Rosenthal and Fode's bias effects. These investigators randomly selected student-experimenters and subjects from all subjects available. They used procedures identical to Rosenthal and Fode's. In addition, prior to the start of the study, each principal investigator at each of the five colleges wrote out his own expectations for the outcome. None of these expectations were confirmed and no significant differences among student-experimenters given high, low, or no expectancy on the person-perception task were found.

Barber and Silver (1968) critically analyzed 31 studies of experimenter bias; 19 of the studies demonstrated bias effects and 12 did not demonstrate bias. After pointing out many inadequate methodological and data-analysis procedures in many of these studies, the authors explored the mediating variables which may be involved.

They first noted that in most studies of experimenter bias not only was an expectation of certain results induced, but also there was an explicit or implicit statement that the experimenters should obtain these results. Therefore, the expectancy was confounded with motivation.

While Rosenthal (1966) divided mediating variables into two types of experimenter error -- intentional and unintentional (see page 9), Barber and Silver separated 11 possible modes of mediation into two sets -- those variables that do not affect the subject's responses and those that do. Their set of variables not affecting the subject's responses was as follows:

1. unintentional misjudgment of the subject's responses;
2. intentional misjudgment of the subject's responses;
3. unintentional misrecording of the subject's responses;
4. intentional misrecording of the subject's responses; and
5. fabrication of the data.

Their second set of variables -- those affecting the subject's responses -- was as follows:

1. unintentional paralinguistic cues (e.g., tone of voice);
2. intentional paralinguistic cues;
3. unintentional kinesic cues;
4. intentional kinesic cues;
5. unintentional verbal reinforcement; and
6. intentional verbal reinforcement.

Barber and Silver concluded that the "experimenter bias effect appears to be more difficult to demonstrate and

less pervasive than was implied in previous reviews." Yet, they did allow that the effect may at times influence experimental results and suggested that results of research not be accepted until replicated by independent research efforts.

One mediating variable which has received little attention is that of the method of inducing the expectancy. Kennedy (1969) conducted two studies on experimenter bias. Both studies used a verbal conditioning task and both had graduate students as experimenters and undergraduate students as subjects. The experimental task required the subjects to make up sentences, each sentence utilizing one of six specific pronouns and a specified verb. The student-experimenters were instructed to respond with "mm-hmm" to all sentences beginning with a self-referrant pronoun. Prior to the start of the experiment, all student-experimenters were given expectancies about whether or not conditioning would occur. In the first study an "indoctrination" program was used. The student-experimenters participated in informal discussions of past research findings, examination of a biased proposal and pilot sessions which were rigged. One-third of the experimenters received a positive expectancy, one-third, a negative expectancy, and one-third, a neutral expectancy; the neutral-expectancy group did not undergo any

indoctrination. Kennedy found no significant biasing effect; i.e., there were no differences between the three groups. Further, this study varied the method of presentation of the stimulus verbs -- either on cards held by the experimenter which gave the experimenter a chance to provide visual cues or by overhead projector with the experimenter seated behind the subject, eliminating the opportunity for visual cues. No significant differences due to mode of stimulus presentation were found.

In his second study, Kennedy informed the student-experimenters that a similar experiment had been conducted recently but that the subjects' performance was quite variable. He then stated that he now could account for the variability by one key factor, that being the subject's philosophical orientation as measured by Wrightman's Philosophy of Human Nature Scale. The student-experimenters were led to believe that a "humanistically-oriented" subject would condition but that a "deterministically-oriented" person would not. All subjects were randomly assigned the labels of "humanistically-oriented" or "deterministically-oriented"; each student-experimenter was assigned five "humanistic" and five "deterministic subjects. The data revealed a significant biasing effect ($p < .02$) -- that is, subjects whom student-experimenters expected would condition actually did condition, while those subjects the

student-experimenters suspected would not condition failed to condition. Kennedy's work not only demonstrates that methods of expectancy induction may be an important variable in experimenter bias, but also reinforces Barber and Silver's view that the experimenter bias effect may not be a pervasive factor which invades all research with human experimenters, but that a complex set of variables operate to produce the effect in some research. In addition, while Kennedy found a significant bias effect in his second study, the effect of the experimenter expectancy accounted for only 6% of the total variance in the data.

In conclusion, several studies have suggested that knowledge of expected outcomes can bias the data an experimenter obtains (Miller, 1972; Rosenthal & Fode, 1963a; Rosenthal & Fode, 1963b);others have found biasing effects under certain conditions and not under others (Dusek, 1971, 1972; Kennedy, 1969; Page, 1972); still others have found no evidence for experimenter bias (Barber et al., 1969). The phenomenon appears to exist, at least under certain conditions, but is more complex than originally thought. According to Kennedy's (1969) data, the phenomenon also accounts for much less of the variability of an experiment than originally hypothesized. Nevertheless, researchers need to be aware of possible bias effects and control for

variables known to influence data, especially intentional error.

Teacher Expectancy and Labelling

Another phenomenon closely related to experimenter bias and also a controversial topic introduced by Rosenthal is that of teacher expectancy. Teacher expectancy may be defined as the holding of certain expectations by the teacher about the behavior and/or academic performance of her students. When these expectations are confirmed -- that is, students behave and/or perform as the teacher expects -- the teacher expectancy has "biased" the data. This confirmation of teacher expectations is often referred to as the self-fulfilling prophecy. The implication is that the teacher somehow conveys her expectation to the pupil (possibly through differential behavior or through the same mediating variables as experimenter bias) and that the pupil's discernment of the differential treatment affects his behavior in the expected direction.

In Pygmalion in the Classroom (1968), Rosenthal and Jacobson discussed the self-fulfilling prophecy which they claim leaves lower-class students at a disadvantage -- that is, teachers "size up" their students on the first day, form expectancies regarding academic achievement based on middle-class values and find their expectations fulfilled at the end of the year. To confirm this theory, the

authors conducted a school-wide research project at "Oak School" in which all students were given the Tests of General Ability (TOGA) at the end of the school year. The following fall, the teachers were given a list of randomly-selected students and told that these students, based on the test results, would "bloom" intellectually during that school year. In May of that year (and the following year), the teachers re-tested all the students. Using pre- to post-test difference IQ scores as the measure of intellectual growth, Rosenthal and Jacobson reported that the experimental group (randomly-selected "bloomers") made significant gains ($p = .02$) over the control group and that these gains were "dramatic" for the first and second grade. This reported phenomenon has led to a greater controversy than Rosenthal's original work on experimenter bias. Reviews of the Pygmalion book (Snow, 1969; Thorndike, 1968) aimed sharp barbs at the authors. As Thorndike pointed out, the obvious errors in the original data cast strong doubts on the conclusions. For example, on the pre-test, both verbal and reasoning IQ scores were obtained. However, the first grade had a mean reasoning IQ of 58; either the test was inaccurate or the population of children was skewed. Snow discussed even further the "serious measurement problems and inadequate data analysis," one of the greatest problems being the failure to use the raw data

in the analysis and the failure to account for initial differences by using an analysis of covariance.

Claiborn (1969) did use a covariance analysis in a similar study and found no significant differences between children labelled "bloomers" and control students when initial individual differences were controlled. However, his pre- and post-tests were only two months apart and his pre-test was administered at the beginning of the spring semester after teachers had spent half of the year with their students and already formed their judgments regarding intellectual ability of individual students. Claiborn also made observations of teacher behavior prior to induction of expectancy and for two weeks after. He found no significant change in the behavior of the teachers toward the "bloomers." Of course, this data is subject to the same criticisms as listed above.

Fleming and Anttonen (1971a, 1971b) also investigated teacher expectancy. Rather than inducing expectancies in the fall, they merely presented teachers with actual IQ scores, no IQ information, inflated IQ scores (+16 points) or Primary Mental Abilities percentiles. They gave teachers no other information and had no further contact with the teachers until post-testing in May; they felt this was a more "natural" situation -- that is, teachers were given the information they normally receive and used the

information in whatever manner they normally did. On the basis of change scores (without accounting for initial differences), these researchers reported no significant differences between groups. In two similar studies, Laskaris (1972) and Palardy (1969), the data are equivocal with regard to teacher expectancy.

Janzen (1972) utilized a different paradigm to study the effects of teacher expectancy. He employed a choice task. The teacher and student were seated at a table opposite one another and divided by a partition; the student could see the teacher's face but not her hands. On the ledge of the partition were three containers. The teacher placed a disc in one of the three containers and the student attempted to select the correct container. During certain experimental conditions, the teacher was required to use a specified order of disc placement, but at other times, she could determine the order herself. The teachers were allowed to reinforce the students for correct choices and could select prizes to reward the children. Janzen found no evidence of differential treatment (reinforcement) of students of different ability levels by teachers.

Antonoplos (1972) examined teacher expectancy with a third paradigm. This paradigm is equivalent to the social psychology experiments of labelling (see below). He had

teachers rate their expectations for children in behavioral descriptions written by the researcher. He found significantly different expectations based on sex and race ($p < .001$) favoring females and whites. This study supports Snow's (1969) conclusion that while Rosenthal and Jacobson's research is questionable, their work has some important social implications -- there may in fact be such a phenomenon which discriminates against minorities.

In the field of social psychology, the similar effect of labelling has been studied. Labelling is the attachment of an ethnic, religious, socio-economic, racial, intellectual, or sexual identification to a person. A biasing effect occurs when a label alters the data -- for example, when a neutral picture or behavioral description produces different reactions (usually measured by rating scales) when given different labels.

Razran (1950) investigated the effect of placing an ethnic label (name) on neutral pictures of female college students. He found significant differences in ratings of the pictures without a name and ratings, two months later, of the same pictures with ethnic names, such as Jewish, Irish, Italian, attached.

Routh and King (1972) assigned occupations indicative of social class to some behavioral descriptions. Both professional clinical psychologists and introductory psychology

students judged the descriptions in reference to the need for professional help. The assignment of these socio-economic labels affected judgments of whether or not the person described needed professional help. Persons with lower socio-economic occupations were more often deemed in need of professional help. Lee (1968) reported similar findings. Sixty psychiatric residents rated a taped interview. Unknown to them, the interview was with a professional actor instructed to act like a normal person without psychological problems; his script contained no indications of socio-economic status. One-third of the residents received no pre-tape information; one-third received pre-listening information indicating high socio-economic status; and the other third was given information indicating low socio-economic status prior to hearing the tape. The residents given the low socio-economic information more often diagnosed the "patient" as mentally ill.

A final example of labelling is Guskin's (1962a, 1962b) work with regard to mentally defective children. He concluded that the label of mental retardation can affect judgments of observers but that this effect depends on the stimulus properties of the particular individual being observed (i.e., physical abnormalities, speech defects, etc.).

The studies reviewed demonstrate that labelling can produce "bias" effects (Antonoplos, 1972; Lee, 1968;

Razran, 1950; Routh & King, 1972). However, all of these studies utilized pictures or behavioral ratings as the stimulus objects. Studies using real persons as stimulus objects (Guskin, 1962a, 1962b) found that the biasing effect was dependent upon the stimulus properties of the individual.

Observer Bias

A second basic phenomenon closely related to experimenter bias is observer bias. Observer bias is defined similarly to that of experimenter bias. Observer bias occurs when observers "bias" their observational data in the direction of the hypothesis. As with experimenter bias, this observer bias is predicated upon knowledge of expected outcome prior to data collection.

Kass and O'Leary (1970) videotaped an elementary class. They had tapes of both baseline and treatment conditions. Observers were trained to use a complex classroom-behavioral code by which to record the behavior of specified children. One-half of the observers were told to expect a decrease in disruptive behavior from baseline to treatment conditions; the rest of the observers were given the opposite expectancy. They reported a significant bias effect among observers but stated that the effect was relatively small. However, the video tapes used showed a dramatic reduction

in disruptive behavior from baseline to treatment so that those observers given an increase expectancy probably found the information discrepant with their observations. Had the change in behavior been less dramatic or if there had been no change in behavior from baseline to treatment, the bias effect may have been greater.

Kent, O'Leary, Diament and Dietz (1972) did indeed use video tapes which showed no change in disruption from baseline to treatment. Two groups of observers were given either a no-change or a decrease expectancy. While global evaluations (on a questionnaire) were affected by the expectancies, the actual data -- the behavioral recordings -- were not. However, O'Leary (1973) has shown that the actual data can be significantly biased if the observer receives feedback from the experimenter regarding the extent to which the observer's data "fits" the hypothesis. In other words, as the observer turns in his data, the experimenter verbally reinforces him for data supportive of the hypothesis and may even "punish" the observer for data unsupportive of the expected outcome (e.g., "This data is not at all in line with our previous research findings"). This effect is similar to the biasing effect of early data returns on experimenters (Rosenthal, 1966).

Another phenomenon related to observer bias is what O'Leary (1973) discussed as observer drift. Kent et al.

(1972) noticed that there were differences in the data of observers given no expectancy, so these differences could not be attributed to treatment conditions. They designed an experiment to investigate the phenomenon and found that groups of observers modify the behavioral code through informal conversations, and that while demonstrating high reliability with each other may not be reliable with other groups of observers. Therefore, the observations may not be valid, according to the original criterion of the behavioral code.

The above findings tend to suggest a significant observer bias effect. On the other hand, Skinrud (1971) found no evidence to support observer bias. He compared the data of "informed" observers (knew whether observations were on a normal or deviant family and whether during baseline or treatment) with the data of a "blind" observer and reported no significant differences. However, only one "blind" observer was used and this observer may have been able to discern when baseline and treatment conditions were occurring, due to differential observational procedures for the two conditions. In addition, Skinrud employed an elaborate coding system which resulted in much variability

between observers thereby increasing the error term and reducing the chance of gaining significance.[1]

In addition to biasing effects with observers, there is the problem of observer reliability. Reid (1970) reported that when observers knew reliability was being assessed, they obtained a mean reliability of .75; when the observers were told that reliability would not be assessed, their reliability dropped to .51. In a similar investigation (Romanczyk, Kent, Diament, & O'Leary, 1973), two reliability assessors were given slightly different variations on four of nine behavioral categories. Observers were told which assessor -- the known assessor -- was checking their reliability during different observation sessions. Reliability with the known assessor was .77, but with the unidentified assessor, was .53. When reliability was assessed covertly (i.e., observers were not aware of the reliability assessment), the mean reliability dropped to .33.

Finally, the reliability of self-observers has been investigated. Self-observers are aware of the target behavior and although not given overt expectations may possess their own expectations and/or value judgments about

---

[1]For a detailed discussion of the effects of an elaborate versus simple behavioral code, see Lipinski and Nelson, 1974a.

that target behavior. Lipinski and Nelson (1974b) have shown self-observers to be unreliable as compared to independent observers; self-observers are especially unreliable when they are unaware of the reliability assessment.

The evidence is fairly strong that knowledge of expected outcome in conjunction with feedback from the experimenter can bias the data recorded by observers (O'Leary, 1973). In addition, the evidence is almost conclusive that knowledge of a reliability assessment can increase observer reliability while knowledge that the observer is not being checked for reliability can decrease reliability (Lipinski & Nelson, 1974b; Reid, 1970; Romanczyk et al., 1973).

Parental Estimates

Parental estimates of the intelligence of their retarded children has been one method of assessing the reliability of parents' global observations of their children. The usual technique for obtaining parental estimates is to ask parents at what age (of a normal child) their child is functioning in specific areas, assuming the estimated age to be a mental age. The ages are then computed into IQ scores (MA/CA X 100). Finally, the child is tested with an intelligence test. Using this method, significant differences were found between parental estimates and actual

tested IQ scores of children in some cases (Ewert & Green, 1957; Heriot & Schmickel, 1967), but not in others (Schopler & Reichler, 1972; Schulman & Stern, 1959; Wolfensberger & Kurtz, 1971). Where a significant difference was found, the parents overestimated the children's ability.

Another procedure for parental estimation of children's intelligence is to separately administer the same intelligence test to both parent and child, instructing the parent to respond to the items as he expects his child to respond. Using this procedure, Gorelick and Sandhu (1967) and Capobianco and Knox (1964) both reported significant differences between parental estimates and actual IQ scores with the parents overestimating the children's ability. However, when Capobianco and Knox analyzed their data by sex of the parent, they found fathers to be accurate estimators but mothers were quite inaccurate.

The second procedure above should be the more valid indicator of parental estimates since in the first procedure, parents only estimate an average age for broad areas of development. However, in the latter method, parents respond to specific tasks.

Parents as Experimenters and Observers

Little research has been done on the reliability of parents recording actual data in the home. Hall et al.

(1972) used parents as both experimenter and observer in four different cases. In all cases, reliability assessments were made. The authors reported 100% agreement between the parent observer and the reliability checker -- other parent, sibling, aunt, neighbor -- in three cases. But the behaviors to be recorded were simplistic and unambiguous: wearing an orthodontist device, completing household chores, dressing time. The one case which involved a frequency count of two behaviors (not an elaborate coding system like those often used by trained observers) had a mean reliability of .855. Another problem with the reported reliabilities was the independence of the observations; often one person verbally reported to the other and in one case, one parent's data was posted in the kitchen in full view of the reliability checker.

Dammann (1973) investigated the ability of parents to accurately judge responses of their child. Twenty-eight families with an adolescent receiving psychiatric services participated in the study. The families were divided into two groups, according to Haley's measure of sequences of speech: high and low communication pathology. The parents were trained through modelling and practice via video tape. Their task was to verbally reinforce positive self-referrant statements (PSR's). All parents had to reach a criterion level of 80% correct reinforcements with the

video tape before proceeding with the identical task with their own child. All parents achieved the criterion level with the video tapes and no significant differences between the high and low groups appeared during practice sessions. However, with their own child, the high communication pathology parents reinforced significantly more negative self-referrants ($p < .05$) than the low communication pathology parents. In addition, the high communication pathology parents more often agreed on reinforcement for non-PSR's ($p < .05$) than the low group. Both groups reinfo ed positive self-referrants with the same accuracy. An her significant finding was that there was no change in the self-referrant statements of the children of either group over time -- i.e., as the interview progressed; this indicated that the parents were not administering effective reinforcement.

One final consideration is the parent as a biased experimenter. Johnson and Lobitz (1972) trained observers to collect data in home settings. Without the observer's knowledge, he instructed the mothers to make their children appear "good" or "bad" on alternate nights. The observers used a complex behavioral code, coding both parent and child behaviors, and a stringent procedure for determining reliability; the obtained mean reliability was .68. Analysis of the data revealed that indeed there was significantly

more deviant behavior on nights the mother was instructed to have her child appear "bad" ($p < .01$) and that on these "bad" nights, there was a significantly higher amount of parental commands ($p < .005$) and negative responses from the parents ($p < .01$). Interviews conducted after the conclusion of the experiment revealed that the mothers were aware of their behavior which led to an increase in the deviant behavior of their children. In Rosenthal's (1966) terminology, the mothers were exhibiting "intentional" experimenter error.

Summary

The work of Rosenthal (1963, 1966) and others (Brager, 1969; Miller, 1972) indicated that experimental results may be biased by the experimenter's knowledge of expected results. However, other research has failed to demonstrate such a bias (Barber et al., 1969) or has found biasing effects under some conditions but not others (Dusek, 1971, 1972; Kennedy, 1969; Page, 1972).

The results regarding teacher expectancy are also equivocal. While Rosenthal and Jacobson (1968) claimed a significant biasing effect, others found no supporting evidence (Claiborn, 1969; Fleming & Anttonen, 1971a, 1971b; Janzen, 1972). There is more support for an effect due to labelling (Antonoplos, 1972; Lee, 1968; Razran, 1950;

Routh & King, 1972); however, this effect appears to be due to the particular stimulus properties of the individual (Guskin, 1962a, 1962b).

Stronger evidence (Kass & O'Leary, 1970; O'Leary, 1973) exists for a bias effect by observers. In addition, observer reliability is reduced when observers know their reliability is not being assessed (Lipinski & Nelson, 1974b; Reid, 1970; Romanczyk et al., 1973).

The parent as observer has not been researched directly However, parental global observations (estimates of child's intelligence) have been assessed. According to some researchers, parents overestimate their child's ability (Capabianco & Knox, 1964; Ewert & Green, 1957; Gorelick & Sandhu, 1967; Heriot & Schmickel, 1967); but other researchers found parents to be accurate estimators (Schopler & Reichler, 1972; Schulman & Stern, 1959; Wolfensberger & Kurtz, 1971). With simple observation procedures, Hall et al (1972) found parents to be accurate observers; the accuracy decreased with a more difficult procedure. Dammann (1973) discovered that parents could accurately judge responses of other children but not their own child. Finally, Johnson and Lobitz (1972) reported that parents can significantly alter the behavior of their children upon "demand" by the experimenter.

CHAPTER III

METHOD

Subjects

Originally eight parents (four mothers and four fathers) who were beginning a parent training program at Henry Wiseman Kendall Center for the Developmentally Disabled served as subjects. One father was unable to complete the study due to personal problems. All parents had a child enrolled in a day-care program at Kendall Center. The parents ranged in age from 24 to 33 years of age; all were within middle-income economic levels. No retarded parents were used as subjects.

Independent Observers

Two independent observers were used to establish a criterion for all video tapes. The two independent observers made their observations separately but were aware of the reliability assessment. For each interval where there was disagreement, the two observers jointly viewed the interval until they agreed on the coding for that interval.

Video Tapes

Ten-minute video tapes were made of the four children whose parents served as subjects. The children were all participants in a day-care program at Kendall Center for

the Developmentally Disabled and had a primary diagnosis of mental retardation (IQ scores ranged from 58 to 63). All the children were male, with ages ranging from 3 years, 11 months to 6 years, 0 months. Video tapes were made of the children while they were in a special education "zone."[2] In this "zone," the children were either working in small groups with the teacher or were involved in individual "academic" (basic discrimination tasks) training with the teacher. The behaviors required of the children in this zone were sitting in chair, attending to the task or to the teacher, and responding to the teacher or task. Most of the disruptive behavior occurring was out-of-chair, modified out-of-chair, non-attending and non-participatory behavior.

Prior to the start of the experiment, much taping of the four children was done. The tapes were observed by a trained observer. From these observations, one 10-minute segment of tape per child was selected so that each of these segments was similar in the number of positive and number of negative behaviors occurring. Similar was defined as 40 ± 2 positive behaviors and 40 ± 2 negative

---

[2]Kendall Center's day-care program revolves around "zones" or treatment areas. The children rotate through the zones every 20 minutes. Zones currently in use at Kendall are special education, speech therapy, self-help training, socialization, physical therapy and individual therapy.

behaviors occurring per 10-minute segment of tape. After selection of the four tapes, the two independent observers observed the tapes to establish the criterion coding and to verify that the tapes were equivalent with respect to the number of positive and negative behaviors. The tapes were then used with the parents.

Behavior Code

When observing the video tapes, independent observers and parent observers used the same behavioral code. Observations were made for 10 seconds followed by 5 seconds of recording, so that there were four observation intervals per minute. The video tapes were edited so that during the recording intervals, the tapes were blank. A "beep" 2 seconds before the observation intervals began served as a stimulus to look up after recording. This editing was done to avoid the distraction (and possible inaccuracy) of each observer (both independent and parent) keeping time on a stop watch.

Independent observers and parent observers both used the same coding sheets to record their observations [See Appendix A]. Each sheet was divided into forty, 10-second intervals; therefore, one sheet accomodated observations of one video tape. Six behaviors were coded on the sheets:

1. Attending/general head orientation
2. Not attending
3. Responding
4. Out-of-chair
5. Touching other's property
6. Touching other's person[3]

The attending and responding behaviors were used as the positive behaviors for the analyses; the not attending, out-of-chair and touching other's property categories were the negative behaviors. Touching another person was a neutral behavior and was included to add complexity to the code. It was only used for the reliability analyses. Positive and negative status of behaviors was assigned on the basis of a questionnaire given to 26 parents of children at Kendall Center. All 26 parents rated "paying attention to the teacher" and "responding to the teacher's question or direction" as "good." Twenty-five of the 26 parents rated "out of chair completely" as "bad." While not included on the parent's questionnaire, touching other's property is considered an inappropriate behavior in many of the behavior codes reported in behavioral research (O'Leary & O'Leary, 1972).

During any interval either attending or not attending had to be recorded; that is, one of the two but not both had to be recorded each interval. The other behaviors might or might not be recorded during a single interval.

---

[3]For a complete description of the behavior code, see Appendix B.

If a behavior was occurring at the end of one interval and was still in process at the beginning of the next interval, it was to be recorded in that next interval -- just as if it were a new behavior.

Procedure

For the introductory session, the parents met as a group. A psychologist from Kendall Center introduced the experimenter and briefly explained that as part of their parent training they would be learning how to make observations with a complex time-sampling procedure. They were instructed to do their best because observation is a skill they would be using with their own children as their training progressed.

Each parent was given a copy of the behavior code [see Appendix B]. The code was read aloud and each category briefly explained and examples demonstrated by the experimenter. Questions were answered. Illustrations of behaviors which would and would not be coded in the particular categories were shown to the parents via video tape. Again, questions were answered. Next, the parents viewed a 2-minute training tape and practiced recording. After they finished viewing the training tape, each individual interval was viewed and the correct coding cited. Finally, to obtain more practice recording on the coding

sheet and within the 5-second period, the parents viewed the training tape a second time and again recorded their observations. Again, the correct coding was cited. The parents were instructed to take the codes home and review them.

For the second session, parents were brought individually into a small therapy room at Kendall Center. The room has a desk, table, chairs and video tape equipment. Prior to observing the experimental tapes, each parent viewed the 2-minute training tape. For each interval which was incorrect, the interval was re-viewed until the parent could correctly pick out the behaviors.

The four experimental video tapes were shown in the same order to all parents. In other words, all parents saw the tapes in order from #1 to #4. A Latin square was used to randomly assign tapes to the four conditions -- neutral, positive expectancy, negative expectancy, own-child. Each video tape was used once for each condition with the mothers and once for each condition with the fathers. Since the tapes were always shown in the same order, the expectancy conditions and own-child condition occurred in random order, thereby evenly distributing any fatigue or practice effects across conditions.

Statements (labels) served to induce expectancies under the positive and negative expectancy conditions.

The neutral and own-child tape received no expectations. The neutral tape served both as the control for the expectancy conditions and as the other-child tape in the analyses involving status of child (own-child versus other-child).

Before observing the positive-expectancy tape, each parent was read the following statement:

> The child on this tape is one of the best children at Kendall. All of the staff members enjoy working with him because he rarely misbehaves. He's a pleasant child and has really made a lot of progress since he's been here.

The following statement was read before the parents observed the negative expectancy tape:

> This tape shows one of the problem-children here at Kendall. He hasn't made any progress since he's been here mostly because he's always misbehaving. He's fussy and hard to work with.

Some specific instructions were read before each of the four tapes [see Appendix C].

Following observations of all four experimental tapes, a Polaroid picture of each child was shown. As each picture was presented (in the same order as the tapes), the parent was asked to rate that child on a short questionnaire [see Appendix D]. This questionnaire was used to determine if the expectancies "took" -- i.e., affected "global" ratings of the children.

Statistical Analysis

As previously mentioned, one male parent was dropped from the study resulting in an unequal number of subjects per group. Therefore, all data were analyzed by an unweighted-means analysis of variance as described by Winer (1971).

CHAPTER IV

RESULTS

Reliability Assessment

The method for calculating reliability was the standard procedure of agreements over agreements plus disagreements (Lipinski & Nelson, 1974b); absences were not included as agreements. Since attending and not attending are exclusive behaviors (i.e., if one occurs, it automatically excludes the other), one reliability assessment was made of these two coded behaviors. That is, for each interval either an agreement (both observers coded A or both observers coded NA) or a disagreement (one observer coded A and one observer coded NA) was recorded. The other four categories being inclusive, each could be recorded as either an agreement (both observers coded the category) or a disagreement (one observer coded the category and the other did not). Therefore, there were a maximum of five possible agreements per interval. The total number of agreements per tape was divided by the total number of agreements plus disagreements per tape. This provided a reliability estimate for each tape for each parent.

Reliability estimates for the tapes ranged from .700 to .852 with the average being .790. A mean reliability was calculated for each parent. These mean reliabilities

are shown in Table 1. A t-test on these overall reliability means revealed no significant differences at the .05 level between the overall mean reliabilities of the mothers ($\bar{X}$ = .791) and the overall mean reliabilities of the fathers ($\bar{X}$ = .783).

To test hypotheses 2 and 3, an unweighted-means analysis of variance, summarized in Table 2, was performed on the reliability estimates for the own-child and other-child tapes. No significant differences appeared due to sex of parent (A) or the own-child vs. other-child conditions (B).

Expectancy Conditions

As previously stated, the four video tapes were edited so that each had 40 ± 2 positive behaviors and 40 ± 2 negative behaviors. To investigate the effect of the expectancy statements (labels), analyses were performed on the number of positive behaviors and negative behaviors recorded by the parents.

To examine hypotheses 4 and 7 and question 5, an unweighted-means analysis of variance (see Table 3) was carried out on the number of positive behaviors recorded for each tape. There were no significant differences due to sex (A) nor expectancy conditions (B). Similarly, hypotheses 5 and 6 and question 5 were tested with an unweighted-means analysis of variance (see Table 4) on the

TABLE 1

Mean Reliabilities For Individual Subjects

Subject	Mean Reliability
S_1	.755
S_2	.822
S_3	.769
S_4	.817
S_5	.780
S_6	.762
S_7	.806

TABLE 2

Sex of Parent X Own-Other Conditions Unweighted-Means Analysis of Variance for Reliability

Source	df	MS	F
Between Subjects			
Sex of Parent (A)	1	.000003	.001250
Subj. w. Groups	5	.002400	
Within Subjects			
Own-Other Conditions (B)	1	.000030	.009677
A X B	1	.000003	.000967
B X Subj. w. Groups	5	.003100	

TABLE 3

Sex of Parent X Expectancy Conditions Unweighted-Means Analysis of Variance for Positive Behaviors

Source	df	MS	F
Between Subjects			
Sex of Parent (A)	1	7.00	.12
Subj. w. Groups	5	60.47	
Within Subjects			
Expectancy Conditions (B)	2	2.05	.09
A X B	2	.33	.02
B X Subj. w. Groups	11	22.73	

TABLE 4

Sex of Parent X Expectancy Conditions Unweighted-Means Analysis of Variance for Negative Behaviors

Source	df	MS	F
Between Subjects			-
Sex of Parent (A)	1	7.68	1.27
Subj. w. Groups	5	6.04	
Within Subjects			
Expectancy Conditions (B)	2	29.36	2.66
A X B	2	14.12	1.28
B X Subj. w. Groups	11	11.02	

number of negative behaviors recorded for each tape. Again, there were no significant differences due to sex (A) nor expectancy conditions (B).

To determine if parents' expectations of their own children would affect their data (hypotheses 8 and 9, question 7), analyses of positive and negative behaviors recorded for the own-child and other-child (neutral) tapes were conducted. As can be seen in Table 5, an unweighted-means analysis of variance on the positive behaviors revealed no significant differences for sex (A) or conditions (B). The unweighted-means analysis of variance on the negative behaviors (see Table 6) produced similar results.

Check on Manipulation

To determine if the parents were affected by the verbal expectancies, the parents responded to four statements on each child [see Appendix D]. An individual analysis was performed on each statement.

The first statement, "This child is very well-behaved," was analyzed by an unweighted-means analysis of variance (see Table 7). There were no significant differences due to sex of parent (A). However, there was a significant difference ($F = 10.37$; $df = 3, 15$; $p < .01$) due to conditions (B). This effect was quite robust ($w^2 = .43$). A

TABLE 5

Sex of Parent X Own-Other Conditions Unweighted-Means Analysis of Variance for Positive Behaviors

Source	df	MS	F
Between Subjects			
Sex of Parent (A)	1	9.054	.226
Subj. w. Groups	5	40.075	
Within Subjects			
Own-Other Conditions (B)	1	2.625	.200
A X B	1	.054	.004
B X Subj. w. Groups	5	13.075	

TABLE 6

Sex of Parent X Own-Other Conditions Unweighted-Means Analysis of Variance for Negative Behaviors

Source	df	MS	F
Between Subjects			
Sex of Parent (A)	1	.15	.01
Subj. w. Groups	5	11.14	
Within Subjects			
Own-Other Conditions (B)	1	16.72	1.05
A X B	1	6.04	.38
B X Subj. w. Groups	5	15.94	

TABLE 7

Sex of Parent X Conditions Unweighted-Means Analysis of Variance for Question #1

Source	df	MS	F
Between Subjects			
Sex of Parent (A)	1	.05	.08
Subj. w. Groups	5	.63	
Within Subjects			
Conditions (B)	3	4.03	10.37**
A X B	3	.41	1.06
B X Subj. w. Groups	15	.39	

**$p < .01$

Newman-Keuls post-hoc test revealed that the negative expectancy condition was significantly different ($p < .01$) from all other conditions (i.e., received more negative ratings). Further, the own-child condition was significantly different from the positive expectancy condition. But the positive expectancy and the neutral condition were not significantly different. Thus, the order of ratings from positive to negative for question #1 was: positive and neutral tapes, followed by own-child, and finally, negative expectancy tape.

The second statement, "This child attends when someone is working with him," was also analyzed with an unweighted-means analysis of variance (see Table 8). No significant differences were found due to sex (A) or conditions (B).

An unweighted-means analysis of variance (see Table 9) was utilized to examine responses to the third statement, "This child would be easy to work with." While there were no significant differences due to sex (A), there were significant differences ($F = 8.28$; $df = 3, 15$; $p < .01$) due to conditions (B). The magnitude of the effect was strong ($w^2 = .40$). A post-hoc analysis (Newman-Keuls) revealed the negative expectancy to be significantly different ($p < .01$) from all other conditions (i.e., rated more negatively). There were no differences among the other three conditions (positive, neutral, and own-child).

TABLE 8

Sex of Parent X Conditions Unweighted-Means Analysis of Variance for Question #2

Source	df	MS	F
Between Subjects			
Sex of Parent (A)	1	.43	.61
Subj. w. Groups	5	.70	
Within Subjects			
Conditions (B)	3	2.32	2.99
A X B	3	.04	.05
B X Subj. w. Groups	15	.78	

TABLE 9

Sex of Parent X Conditions Unweighted-Means Analysis of Variance for Question #3

Source	df	MS	F
Between Subjects			
Sex of Parent (A)	1	.76	.62
Subj. w. Groups	5	1.23	
Within Subjects			
Conditions (B)	3	5.15	8.28**
A X B	3	.96	1.54
B X Subj. w. Groups	15	.62	

**$p < .01$

The final statement was "I would enjoy working with this child." An unweighted-means analysis of variance, summarized in Table 10, showed no significant differences due to sex of parent (A), but found significant differences due to conditions ($p < .05$). The conditions factor accounted for 26% of the variability of the data. Again, the Newman-Keuls procedure was employed as the post-hoc test of individual differences and again the negative expectancy condition was significantly different ($p < .05$) from all conditions. The positive expectancy, neutral, and own-child tapes were not rated significantly different from each other.

Order Effects

To determine if there were any practice or fatigue effects, an unweighted-means analysis of variance was performed on the reliability estimates for the tapes by order of presentation (i.e., Tape #1, #2, #3, #4). No significant differences were found (see Table 11). Therefore, it was assumed that reliabilities of the parents neither increased or decreased as a result of practice or fatigue. In addition, it appeared that no particular tape was easier to judge and thereby produce higher reliability.

TABLE 10

Sex of Parent X Conditions Unweighted-Means Analysis of Variance for Question #4

Source	df	MS	F
Between Subjects			
Sex of Parent (A)	1	.50	.58
Subj. w. Groups	5	.87	
Within Subjects			
Conditions (B)	3	2.10	5.80*
A X B	3	.10	.22
B X Subj. w. Groups	15	.44	

*$p < .05$

TABLE 11 -

Sex of Parent X Tape-Order Unweighted-Means Analysis of Variance for Reliability

Source	df	MS	F
Between Subjects -			
Sex of Parent (A)	1	.0004	.1275
Subj. w. Groups	5	.0035	
Within Subjects			
Tape-Order (B)	3	.0001	.7936
A X B	3	.0009	.5323
B X Subj. w. Groups	15	.0018	

CHAPTER V

DISCUSSION

As previously mentioned, in recent years behavior therapists have turned their attention from clinical settings to "natural" environments (e.g., Baer & Wolf, 1967; Patterson et al., 1967). In so doing, parents, as mediators within the "natural" setting, often become both observers and experimenters (Hall et al., 1970, 1972). In their capacity as observers it is crucial that high reliability be obtained so that the therapist can conclude that the data accurately reflect the behavior occurring.

In the present study, the parents produced a mean reliability of .79 with less than two hours of training. Considering the amount of training, this reliability is high in comparison with other studies. Patterson and Harris (1968) took six to nine 1-hour training sessions to "bring all observers up to 80% agreement with the trainer." They were training both parents and independent observers. Using a 30-category behavioral code, Skinrud (1971) investigated the effects of observer bias. For his research, he required that observers maintain a .75 minimum level of agreement. This is the same level of agreement that Reid's (1970) observers maintained while they were under an overt assessment condition. However, when the observers did not know reliability was being assessed, their reliability

dropped to .51. Romanczyk et al., (1973) described .52 as a "moderate" level of reliability and .81-.84 as "highly reliable."

Lipinski and Nelson (1974a) have discussed some of the problems and issues related to an acceptable level of reliability. One point they made is that the level of reliability will vary as a function of the coding system -- i.e., the number of categories being recorded. Fewer categories should produce higher reliability because the observers are more likely to agree by chance alone. Another factor they noted was the inclusion or exclusion of absences as agreements. One obtains higher reliability when absences are counted as agreements. Repp, Deitz, Boles, Deitz, and Repp (1974) have substantiated Lipinski and Nelson's (1974a) assertations. They investigated the effects of method of calculation on reliability of observational data. They found that reliability could vary 64% to 94%, dependent on the method of computing the reliability and the interval size. They also found that methods counting absences of recording as agreements produced consistently higher reliabilities. In the present study, if absences were counted as agreements, the parents would have a mean reliability of .91, a difference of 12 percentage points. Considering that these parents received less than two hours of training, that the code included six categories,

and that a stringent procedure for computing reliability was employed, .79 appears to be fairly high reliability.

This high level of reliability with such a small amount of training has important implications for the clinical field. Since behavior is situation-specific, it is essential that therapists train mediators to work with a child in the particular setting in which the problem behaviors occur. However, the gain of working in the "natural" environment may be offset if the training of mediators becomes lengthy. This present research demonstrated that parents can be trained in a short amount of time to serve as reliable observers.

Not only were these parents highly reliable, but their accuracy was unaffected by observing their own child. Contrary to Dammann's (1973) findings, these parents were able to objectively discriminate behavior of their own children equally well as other children. In Dammann's (1973) research, parents of children with psychiatric problems were able to judge responses of other children but not their own children. In light of the present findings, one would assume that there is nothing intrinsic to the parent-child relationship which inhibits objective judgment. Perhaps the inability of Dammann's parents to objectively discriminate responses of their own children was somehow related to the "psychiatric" problems of the children.

So far, the parents in the present study have demonstrated observer reliability plus objectivity in viewing their own children. Despite the claims of experimenter and observer bias by some researchers (Rosenthal, 1963, 1966; Rosenthal & Fode, 1963a, 1963b; Kass & O'Leary, 1970), the data of these parents were not affected by expectancies. The parents were given a positive expectancy, a negative expectancy, and no expectancy (neutral) in addition to viewing their own child. However, there were no significant differences in number of positive or negative behaviors recorded for any of the conditions. This result supports those authors who found no experimenter bias (Barber et al., 1969) and no teacher expectancy effect (Claiborn, 1969; Fleming & Antonnen, 1971a, 1971b).

While their observational data were not affected, the parents global ratings of the children were affected by the expectancies. There were significant differences for three of the four statements on which the children were rated. The one statement which did not produce significant differences was "This child attends..." This statement was probably rated more objectively since it was more related to the code (i.e., attending/not attending) and/or was stated quite specifically. The researchers at The State University of New York at Stony Brook reported similar findings -- that is, no effect on the observational data, but an effect on

effect on the global ratings. While Kass and O'Leary (1970) found a bias effect, their tapes were not neutral. That is, the tapes actually showed a dramatic decrease in disruptive behavior from baseline to treatment. Kent et al. (1972) used equivalent tapes and found no biasing effect regarding the observational data. Yet, they did find that global ratings were affected by the expectancies given. In some of the social psychology and teacher expectancy research which reported expectancy effects due to labelling (Antonoplos, 1972; Razran, 1950; Routh & King, 1972), the differences are based on global ratings. For instance, in Antonoplos' (1972) study, teachers were given written descriptions of children and made global ratings of the children on the basis of the written descriptions. Further, Miller (1972) asserted that the amount of experimenter bias was directly related to the ambiguity of the task. Brager (1969) reported supportive data in that of three tasks, the objective task was not biased while two subjective tasks were biased by expectancies. In the present research, the objective task (observing) was not susceptible to bias while the more subjective or ambiguous task (rating) was. From these present findings and the past research, it appears that global ratings are easily affected by labelling or expectations but that actual observational data is more resistant to expectancy effects.

The implication of the findings reported here is that parents can be trained to be reliable observers with a minimal amount of training. Parents can be objective observers, not allowing their expectations to bias their data. However, their global statements are affected by their expectations. Specific statements (i.e., more behaviorally-phrased or more related to the specific observations) appear to be immune to this bias effect. These results have implications for therapists working with parents. For example, a parent may have a negative set about a child and describe the child in such global terms as "bad," "mean," "hard to get along with," etc. During the interview, the therapist should ask very specific questions (e.g., does he play more by himself or with other children) rather than global questions (e.g., does he get along with other children). Further, the therapist may want the mother to make observations of her child and then critically look at the observations. When examining the observations, again remarks should be specific rather than global. After looking at the observations in specific terms, the global questions can be asked. If negative expectancies still exist, the discrepancy of the negative set with the behavioral data can be explored.

In conclusion, it appears that parents can be reliable, unbiased observers. This conclusion supports the move into

the "natural" environment. While parents appear to be unbiased observers, therapists should be careful not to induce "bias" through reinforcement of particular data. O'Leary (1973) and Lipinski (1974) have demonstrated that feedback (especially reinforcement) can bias data. Further, the present study utilized video tapes and a specific behavioral code. In addition, the parents were aware of the reliability assessment. Further studies should be conducted with parents as observers of on-going behavior in "natural" settings with both overt and covert reliability assessments. Also, other behavioral codes might be utilized, as well as frequency counts and duration measures.

BIBLIOGRAPHY

Antonoplos, D. P. Interactions of teacher-pupil sex as expressed by teacher expectations, patterns of reinforcement and judgments about pupils: A national study. Dissertation Abstracts International, 1972, 32, 6117.

Baer, D. M. and Wolf, M. M. The entry into natural communities of reinforcement. Paper presented at APA, Washington, D. C., 1967. Reprinted in R. Ulrich, T. Stachnik and J. Mabry. Control of Human Behavior. Glenview, Illinois: Scott, Foresman & Co., 1970.

Barber, T. X., Forgione, A., Chaves, J. F., Calverly, D. S., McPeake, J. D., and Bowen, B. Five attempts to replicate the experimenter bias effect. Journal of Consulting and Clinical Psychology, 1969, 33, 1-6.

Barber, T. X. and Silver, M. J. Fact, fiction and the experimenter bias effect. Psychological Bulletin Monograph Supplement, 1968, 70, 1-29.

Brager, G. L. A study of the phenomenon of experimenter bias using three types of classroom tasks. Dissertation Abstracts International, 1969, 29, 4316-4317.

Capobianco, R. J. and Knox, S. IQ estimates and the index of marital integration. American Journal of Mental Deficiency, 1964, 68, 718-721.

Claiborn, W. L. Expectancy effects in the classroom: A failure to replicate. Journal of Educational Psychology, 1969, 60, 377-383.

Dammann, Carrell. Patterns of family interaction and the ability of parents to administer accurate reinforcement. Paper presented at the Southeastern Psychological Association, New Orleans, April, 1973.

Dusek, J. B. Experimenter bias in performance of children at a simple motor task. Developmental Psychology, 1971, 4, 55-62.

Dusek, J. B. Experimenter bias effects on the simple motor task performance of low- and high-test anxious boys and girls. Psychological Reports, 1972, 30, 107-114.

Ewert, J. C. and Green, M. W. Conditions associated with the mother's estimate of the ability of her retarded child. American Journal of Mental Deficiency, 1957, 62, 521-533.

Fleming, E. S. and Anttonen, R. G. Teacher expectancy examined at different ability levels. Journal of Special Education, 1971, 5, 127-131.(a)

Fleming, E. S. and Anttonen, R. G. Teacher expectancy or My Fair Lady. American Educational Research Journal, 1971, 8, 241-252.(b)

Gorelick, M. C. and Sandhu, M. Parent perception of retarded child's intelligence. Personnel and Guidance Journal, 1967, 46, 382-384.

Guskin, S. L. The influence of labelling upon the perception of subnormality in mentally defective children. American Journal of Mental Deficiency, 1962, 67, 402-406.(a)

Guskin, S. L. The perception of subnormality in mentally defective children. American Journal of Mental Deficiency, 1962, 67, 53-60.(b)

Hall, R. V., Axelrod, J., Tyler, L., Grief, E., Jones, F. C., and Robertson, R. Modification of behavior problems in the home with a parent as observer and experimenter. Journal of Applied Behavior Analysis, 1972, 5, 53-64.

Hall, R. V., Cristler, C., Cranston, S. S., and Tucker, B. Teachers and parents as researchers using multiple baseline designs. Journal of Applied Behavior Analysis, 1970, 3, 247-255.

Heriot, J. T. and Schmickel, C. A. Maternal estimate of IQ in children evaluated for learning potential. American Journal of Mental Deficiency, 1967, 71, 920-924.

Janzen, G. G. An experimental study of the differential response patterns of elementary school teachers toward their students. Dissertation Abstracts International, 1972, 32, 6204.

Johnson, S. M. and Lobitz, G. K. Demand characteristics in naturalistic observation. Unpublished, 1972.

Kass, R. E. and O'Leary, K. D. The effects of observer bias in field-experimental settings. Paper presented at a symposium, "Behavior Analysis in Education," University of Kansas, April, 1970.

Kennedy, J. J. Experimenter bias in behavioral research. Unpublished, 1969.

Kent, R. N., O'Leary, K. D., Diament, C., and Dietz, A. Expectation biases in observational evaluation of therapeutic change. Unpublished, 1972.

Laskaris, J. B. Some effects of the experimental manipulation of teacher expectations on the measured academic growth of fifth-grade children. Dissertation Abstracts International, 1972, 32, 6209.

Lee, S. D. Social class bias in the diagnosis of mental illness. Dissertation Abstracts International, 1968, 28, 4758-4759.

Lipinski, D. P. The effects of differential reinforcement on the reliability and reactivity of self-recorders. Unpublished doctoral dissertation, University of North Carolina at Greensboro, 1974.

Lipinski, D. P. and Nelson, R. Problems in the use of naturalistic observations as a means of behavioral assessment. Behavior Therapy, 1974, 5, 341-351.(a)

Lipinski, D. P. and Nelson, R. The reactivity and unreliability of self-recording. Journal of Consulting and Clinical Psychology, 1974, 42, 118-123.(b)

Miller, P. N. The effect of experimenter status and experimenter competency on the magnitude of the experimenter bias effect. Dissertation Abstracts International, 1972, 32, 7294-7295.

O'Leary, K. D. The effects of observer bias in field-experimental settings. Unpublished, 1973.

O'Leary, K. D. and O'Leary, S. G. Classroom management: The successful use of behavior modification. New York: Pergamon Press, 1972.

Page, S. Interpersonal expectancy effects: Studies with three psychiatric samples. Psychological Reports, 1972, 30, 139-142.

Palardy, J. M. What teachers believe -- What children achieve. The Elementary School Journal, 1969, 69, 370-374.

Patterson, G. R. and Harris, A. Some methodological considerations for observation procedures. Paper presented at the American Psychological Association Annual Convention, San Francisco, 1968.

Patterson, G. R., McNeal, S., Hawkins, N., and Phelps, R. Reprogramming the social environment. Journal of Child Psychology and Psychiatry, 1967, 8, 181-195.

Ray, R. S. Parents and teachers as therapeutic agents in behavior modification. Paper presented at the Second Annual Behavior Modification Institute, Tuscaloosa, Alabama, May, 1969.

Razran, G. Ethnic dislikes and stereotypes: A laboratory study. Journal of Abnormal and Social Psychology, 1950, 45, 7-27.

Reid, J. B. Reliability assessment of observation data: A possible methodological problem. Child Development. 1970, 41, 1143-1150.

Repp, A. C., Deitz, D. D., Boles, S. M., Deitz, S. M., and Repp, C. F. Differences among common methods for calculating interobserver agreement in applied behavioral studies. Paper presented at the Southeastern Psychological Association Annual Meeting, Hollywood, Florida, May, 1974.

Romanczyk, R. G., Kent, R. N., Diament, C., and O'Leary, K. D. Measuring the reliability of observation data: A reactive process. Journal of Applied Behavior Analysis, 1973, 6, 175-184.

Rosenthal, R. On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. American Scientist, 1963, 51, 268-283.

Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.

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

Rosenthal, R. and Fode, K. L. Three experiments in experimenter bias. Psychological Reports, 1963, 12, 491-511.(b)

Rosenthal, R. and Jacobson, L. Pygmalion in the classroom. New York: Holt, Rinehart & Winston, 1968.

Routh, D. K. and King, K. M. Social class bias in clinical judgment. Journal of Consulting and Clinical Psychology, 1972, 38, 202-207.

Schopler, E. and Reichler, R. J. How well do parents understand their own psychotic child? Journal of Autism and Childhood Schizophrenia, 1972, 2, 387-400.

Schulman, J. L. and Stern, S. Parents' estimate of the intelligence of retarded children. American Journal of Mental Deficiency, 1959, 63, 696-698.

Skinrud, K. A preliminary evaluation of observer bias in multi-variate field studies of social interaction. Unpublished, 1971.

Snow, R. E. Unfinished pygmalion. Contemporary Psychology, 1969, 14, 197-199.

Stuart, R. B. Behavioral contracting within the families of delinquents. Journal of Behavior Therapy and Experimental Psychiatry, 1971, 2, 1-11.

Tharp, R. G. and Wetzel, R. J. Behavior modification in the natural environment. New York: Academic Press, 1969.

Thorndike, R. L. Review of R. Rosenthal and L. Jacobson, Pygmalion in the Classroom. American Educational Research Journal, 1968, 5, 708-711.

Wahler, R. G. Peers as classroom behavior modifiers. Paper read at the American Association on Mental Deficiency, Washington, D. C., May, 1970.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.

Wolfensberger, W. and Kurtz, R. A. Measurement of parents' perceptions of their children's development. Genetic Psychology Monographs, 1971, 83, 3-92.

APPENDIX A

Observation Sheet

Observer's Name ______________________________

Child ______________ Date ______________

A	NA	A	NA	A	NA	A	NA	A	NA	A	NA
R	OC	R	OC	R	OC	R	OC	R	OC	R	OC
TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr

A	NA	A	NA	A	NA	A	NA	A	NA	A	NA
R	OC	R	OC	R	OC	R	OC	R	OC	R	OC
TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr

A	NA	A	NA	A	NA	A	NA	A	NA	A	NA
R	OC	R	OC	R	OC	R	OC	R	OC	R	OC
TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr

A	NA	A	NA	A	NA	A	NA	A	NA	A	NA
R	OC	R	OC	R	OC	R	OC	R	OC	R	OC
TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr	TPe	TPr

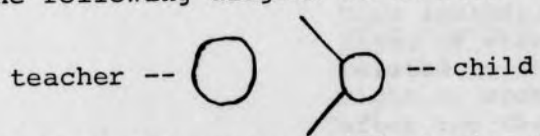
APPENDIX B

Behavior Code

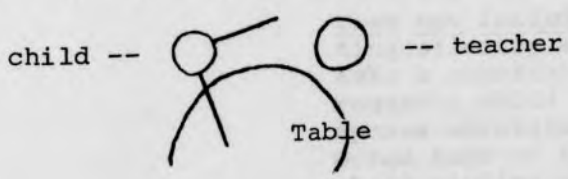
Behavior	Symbol	Description
Attending/ general head orientation	A	Child's head is in a position within a 90° angle -- if the child is facing the teacher, the 90° angle is bounded by lines halfway to the shoulders on either side of a direct frontal position[1]; if the child is facing the table, the 90° angle is bounded by one line at the direct frontal position and one line at the shoulder nearest the teacher.[2] This position must be maintained for the full 10-second interval. Exception: If the teacher is working with the target child and she moves the materials with which they are working out of the angle and the child attends to the materials, then the child is attending and should be coded A.

---

[1]The following diagram illustrates this 90° angle:



[2]The following diagram illustrates this 90° angle:



Behavior	Symbol	Description
		Does include: looking straight ahead with eyes partially or completely closed. Does not include: looking directly up (at the ceiling) or down (at the floor or his chair or the table) even for a second; placing hands over eyes.
Not attending	NA	Any head position for any amount of time outside the 90° angle described above. That is, if the child looks up at the ceiling or around his shoulder for a second, he is not attending.
Responding	R	Responding to the teacher's question, direction or prompt -- the child gives an answer when the teacher asks a question of him, prompts him, or asks a question of the group as a whole; the child does as the teacher directs or makes an attempt to follow the teacher's directions. Does include: any answer, right or wrong; any response related to the directed task, right or wrong; a response after the teacher prompts the answer. Does not include: child sitting silently after a teacher asks a question or makes a request; child making a response unrelated to the directed task -- for example, the child picking up something on the floor when the teacher has requested him to point to a banana.

Behavior	Symbol	Description
Out of Chair	OC	Child is completely out of his seat; child's buttocks is not on seat of the chair. Does include: one foot on floor and one foot on chair; standing in the chair; standing beside the chair with hand(s) on chair. Does not include: child sitting sideways in chair; child sitting on the edge of his chair.
Touching other's property	TPr	Child comes into contact with another's property without permission to do so. Does include: playing with or touching the materials (instructional) used by the teacher without her specific direction to do so; grabbing an object from someone else; touching someone else's chair. Does not include: touching own chair or the table; touching another person or another person's clothing; manipulation of the instructional materials upon the teacher's request.
Touching another person	TPe	The child comes into contact with another person or another person's clothing. Child initiates contact. Does include: hugs, kisses, handshakes, pats on the back, hitting, tugging at another's clothes, arm around another person.

Behavior	Symbol	Description
		Does not include: teacher hugging the child or patting the child. Child must initiate the contact.

APPENDIX C

Special Instructions Regarding Tapes

Tape #1:

The teacher will be out of the room during part of this tape -- an observer will be seated by the table -- she does not count as the teacher. Therefore, he cannot make any responses while the teacher is out of the room. He can be Attending or Not Attending to the teacher's original position (empty chair); he can also be OC, TPe, TPr. Assume he has no permission to touch any property -- the observer can't grant permission.

Tape #2:

The teacher will be out of the room during part of this tape -- he cannot make any responses while the teacher is out of the room. He can be Attending or Not Attending to the teacher's original position (empty chair); when the teacher leaves the area, assume that he does not have permission to touch anything on the table or in the basket by the table unless you see the teacher explicitly hand him something before she leaves.

Tape #3:

(No special instructions given for this tape)

Tape #4:

When the child on this tape goes out of view of the camera, you know he is out of chair, assume he is also Not Attending. Also, remember touching someone else's chair is TPr.

APPENDIX D

Questionnaire for Parent-Observers

Please rate this child on a scale from 1 to 5 as follows:

1	2	3	4	5
Strongly Agree	Agree	Don't Know	Disagree	Strongly Disagree

1. This child is very well-behaved.

1 2 3 4 5

2. This child attends when someone is working with him.

1 2 3 4 5

3. This child would be easy to work with.

1 2 3 4 5

4. I would enjoy working with this child.

1 2 3 4 5