DOCUMENT RESUME

ED 078 086 TM 002 905

AUTHOR O'Leary, K. Daniel

TITLE The Effects of Observer Bias in Field-Experimental Settings. Final Report.

INSTITUTION State Univ. of New York, Stony Brook. Dept. of Psychology.

SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C. Regional Research Program.
PUB DATE Mar 73
GRANT OEG-2-71-0017
NOTE 23p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Attitudes; Behavior Patterns; Children; *Data Collection; Evaluation Criteria; Expectation; *Field Studies; *Hypothesis Testing; *Observation; Questionnaires; Research Methodology; Tape Recordings; Technical Reports

ABSTRACT

A series of studies on observer biases revealed that simply informing observers of experimental hypotheses does not produce observational data consonant with those hypotheses. However, questionnaire responses following an experiment with different induced expectations does produce global data consonant with experimental hypotheses. In addition, if the observers are informed of the experimental hypotheses and the investigator provides daily feedback to the observers indicating how well their data support his hypotheses, the observers will report data consonant with those hypotheses. The method of investigation in the studies reported involved having observers watch specially prepared video tapes of children who exhibited significant amounts of disruptive behavior. Following a pre-treatment or baseline period, observers were then asked to watch video tapes on which children displayed no change or marked reductions in disruptive behavior during a "treatment period." While observer biases per se did not result in confounded data in any of the studies, an unanticipated problem of observer drift or changing observational criteria can result in seriously confounded data where groups of observers initially trained together are later assigned to different treatment conditions. (Author/CK)

Final Report

Grant No. OEG-2-71-0017

The Effects of Observer Bias in Field-Experimental Settings

K. Daniel O'Leary

Psychology Department
State University of New York
Stony Brook, New York 11790

March, 1973

The research reported herein was performed pursuant to Grant OEG-2-71-0017 from the U.S. Office of Education, Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily reflect official office of Education position or policy.

U.S. Department of Health, Education and Welfare
Office of Education
Regional Research Program

Table of Contents

Abstract

A series of studies on observer biases revealed that simply informing observers of experimental hypotheses does not produce observational data consonant with those hypotheses. However, questionnaire responses following an experiment with different induced expectations does produce global data consonant with experimental hypotheses. In addition, if the observers are informed of the experimental hypotheses and the investigator provides daily feedback to the observers indicating how well their data support his hypotheses, the observers will report data consonant with those hypotheses. The method of investigation in the studies reported involved having observers watch specially prepared video tapes of children who exhibited significant amounts of disruptive behavior. Following a pre-treatment or baseline period, observers were then asked to watch video tapes on which children displayed no change or marked reductions in disruptive behavior during a "treatment period." While observer biases per se did not result in confounded data in any of the studies, an unanticipated problem of observer drift or changing observational criteria can result in seriously confounded data where groups of observers initially trained together are later assigned to different treatment conditions. Similarly, experimenters can inadvertently shape data consonant with their experimental hypotheses where they inform the observers of the expected treatment outcome and give them feedback regarding how well their data conform with that expectation.

Introduction

The importance of expectancies of observers and experimenters was sprung upon psychology in bold fashion by Rosenthal in 1963, and since that time expectancy studies have assumed a position of disputed prominence in various circles. In a prototypic experiment (Rosenthal & Fode, 1963), naive rats were randomly assigned to two groups of experimenters in a maze learning study. One group of experimenters (undergraduates) were told that they were testing maze-bright animals and the other group of experimenters were told that they were testing maze-dull animals. Experimenters who were told that their animals were bright reported faster learning times for their animals than the experimenters who were told that their animals were maze dull. Rosenthal extended a variant of this work to classroom settings where he informed teachers that certain randomly selected children in their class were spurters (i.e. "late bloomers" with unrealized academic potential). On the basis of pre- and post-testing in the fall and spring it was found that children in the experimental group, i.e., spurters, had a greater increase in IQ than did the controls (Rosenthal & Jacobsen, 1966). Part of the heat generated from the Rosenthal studies is of course due to the possibility that all psychological experimentation involving an informed human observer could be confounded by the observer bias. However, another equally important reason for such heat is due to the failure of many people to replicate Rosenthal's work in both the laboratory and the classroom (Barber & Silver, 1968; Clairborn, 1969). Despite the failures to replicate and extensive criticisms of Rosenthal' methodology (Snow, 1968; Thorndike, 1968), investigators applying learning principles to the modification of behavior have taken note of the expectancy phenomenon. For example, Thomas, Becker, and Armstrong (1968) noted that their observers were not informed of changes in experimental conditions. McKenzie, Clark, Wolf, Kothera, and Benson (1968) noted that their observers were not informed of the type of home based token procedure utilized in their study nor were they informed when the token program was put into effect. In order to "control for any bias in ratings," O'Connor (1968) kept observers unaware of assignment of subjects to various treatment or control conditions and each observer watched a random combination of treated and control subjects. Bushell, Wrobel, and Michaelis (1968) had classroom observers record behavior descriptions which were later coded as study or non-study behavior. As they stated, "A description might have been coded Non-Study on Day 15 and 'Study" on Day 19 simply because the observer expected study behavior to increase during the final contingent (or treatment phase)." Consequently, Bushell et. al., trained new coders who had no knowledge of the details of the original investigations. Despite these precautions and the more

systematic coding check of Bushell et al., there has been only one study dealing with the effects of observer bias in the classroom studies where the biases or expectations were independently manipulated (Kass & O'Leary, 1970). As a side issue in a study several years ago Scott, Burton, and Yarrow (1967) did find a significant difference in the observations of one informed observer and a group of uninformed observers--using positive and negative acts as dependent measures. That is, the informed observer's records were more confirming of the experimental hypothesis than were those of the uninformed observers. However, the Scott, Burton, and Yarrow study used only one informed observer (the senior author) and the dependent measures were rather global; e.g., positive acts included suggestions, sharing ideas, helping, showing concern for others, and carrying on friendly conversation. In addition, many of the subcategories were quite unreliable with reliabilities ranging from -.09 to .80. More importantly, even though experimenters may not inform their observers of the informal hypotheses, the observers may easily become aware of any experimental changes For example, in a study by Madsen, Becker, and Thomas (1968) where teachers were told to praise appropriate behavior and ignore disruptive behavior, the observers were not informed of the experimental conditions, but the investigators noted that the changes were often dramatic enough that observer comments clearly reflected programmed changes in the teacher's behavior. Furthermore, when a treatment condition is in effect, the experimenter or graduate assistant may subtly or even overtly reinforce the observer for bringing him "good" or confirming data with comments like "That's really interesting," "That teacher is having some effect on those kids,"--or more openly--"That treatment almost never fails to produce an increase in appropriate behavior." Because of the principal investigator's involvement with token reinforcement studies (O'Leary & Becker, 1967; O'Leary, Becker, Evans, & Saudargas, 1969), and the near impossibility of deceiving the observers about the onset and intended experimental effects of a token program the problem of observer bias has been a particularly pressing problem in his research and that of others similarly aware of the expectancy problem and despite admonitions to them to carefully monitor their own behavior in this regard, several observers have reported that although they were aware of the problem, their results might still be biased because of their knowledge of the hypotheses of the study.

Kass and O'Leary (1970) systematically manipulated predictions of treatment effects for three groups of observers who recorded the behavior of children from videotapes of a classroom setting. Two groups of observers were told, respectively, that level of disruptive behavior from "baseline" to treatment" phases of the study a) would increase

and b) would decrease. The third group was given no prediction or results. In fact, all groups of observers viewed the same video tapes which were selected, on the basis of a priori ratings, to show a substantial decrease from baseline to treatment. Significant effects associated with the main treatment manipulation were obtained among the three groups on five out of nine categories. Because of differences among the three groups found during baseline, an analysis of covariance was conducted. This analysis revealed that after adjusting for initial differences during baseline, the three groups still showed significant effects on four of the five categories. Visual inspection of the ordering of the three groups means on these four categories revealed that on two of the four the differences clearly were not in line with the predictions. As will be seen later, the differences obtained in this study may have been largely due to observer drift--or random fluctuations in the observational criteria within the three groups.

Procedures and Results

A. The Effects of Observer Bias

A doctoral dissertation was designed by Kent, 1972, to assess the effects of knowledge of predicted results by observers on behavioral recordings generated under circumstances similar to field-experimental investigations in behavior modification. The experimental variables were: predicted behavior change from baseline to treatment condition (decrease vs. no change); actual behavior change (decrease vs. no change); and expectation induction (prior to baseline vs. subsequent to baseline). An observational code developed to measure the disruptive behavior of children in a classroom was employed. The categories of behavior comprising this code were: out of chair, modified out of chair, touching other's property, vocalization, playing, orienting, noise, aggression and time off task.

Forty observers were trained as a group for seventeen sessions before assignment to the eight experimental groups for three additional training sessions. All groups viewed the same "pre-baseline" and "baseline" videotapes, followed by "treatment" videotapes demonstrating either decrease or no change from baseline levels of disruptive behavior. Eight "baseline" and eight "treatment" ratings were obtained on each of two target children from the five observers in each experimental group.

An analysis of behavioral recordings of "pre-baseline" videotapes revealed greater than five per cent significant differences among the experimental groups on the nine behavioral categories prior to the experimental manipulations. Analysis of "pre-baseline" and "baseline" recordings of four groups which received no experimental intervention until immediately prior to viewing the "treatment" videotapes revealed: a) greater than five per cent significant differences among the experimental groups; b) a greater number of significant differences among "baseline" than "pre-baseline" recordings and c) virtually no similaaity between particular differences which existed in "pre-baseline" and "baseline" recordings. Under this circumstance, the differences in behavioral recordings as a function of the experimental manipulation were completely and inextricably confounded with differences which evolved spontaneously among the experimental groups. That is, groups of observers tend to "drift" or randomly modify their definitions of the behavioral code.

The problem of observer drift was completely unanticipated and this drift may have accounted for the differences attributed to expectation in the Kass and O'Leary, 1970, study. The groups of observers in the Kass and O'Leary stud

were trained separately and were later assigned to separate expectation groups. In fact, the pretreatment differences of Kass and O'Leary (1970) and Kent (1972) are now clearly attributable to this observational drift.

Because of the problem of observer drift obtained, Kent, O'Leary, Diament, and Dietz (1972) designed a study to re-examine the effects of predicted results (expectation) on the observational recordings of trained observers. This study was specially designed to avoid the possibility that differential "drift" in definition of the behavior code in the experimental conditions would be confounded with the effects of predicted results. Videotapes of children in a classroom during "baseline" and "treatment" conditions were rated by two groups of observers, employing a standard nine category behavioral code for disruptive behavior. The two groups of 10 observers were told that they were viewing, respectively, the effects of a) a token program which would dramatically reduce the level of disruptive behavior from baseline and b) a control program which would produce no change from baseline. In fact, the same videotapes were viewed by all observers. A priori ratings of a pool of videotapes were utilized to create "baseline" and "treatment" conditions which were matched for level of disruptive behavior. However, after each "treatment" recording period, observers in each group were told that a casual examination of their recordings indicated that the predicted results were emerging. This was intended to increase the similarity to field settings in which such casual feedback may often be given, and to enhance the liklihood that biases due to predicted results would occur. This design used, within each experimental condition, [illegible] pairs of observers who computed reliability only with [illegible] pairs. Thus "drift" among the five pairs of observers who were told they would view the effects of a token procedure, and among the five pairs of observers who were told they would view a control procedure, could be separated from the effects of predicted results on behavioral recordings. Following the final experimental session, both groups were given a questionnaire to determine whether observers understood the results predicted for their group. In addition, observers were asked what they anticipated and what they perceived as the results of the experimental condition they viewed. Finally, all observers were asked if they felt they had been misinformed about any aspect of the study.

Global evaluations of treatment effects obtained on a post questionnaire were significantly affected by predicted results but behavioral recordings were not. That is, although there were no differences in the actual frequencies of behavior recorded in the two experimental groups, when observers were asked "What actually happened to the level of disruptive behavior from the baseline to the treatment

condition," they reported data consonant with the experimental hypotheses. Nine of the ten observers for whom a decrease in level of disruptive behavior from baseline to treatment conditions was predicted reported actually viewing a decrease. Seven of the ten observers for whom no change was predicted reported viewing no change.

While no observational differences were obtained in the Kent et. al. (1972) study that could be attributed to induced expectations, it was still possible that induced expectations combined with experimenter feedback indicating how well the observational data fit with his predictions would result in biased data. Consequently, a study designed to shape data consonant with experimenter hypotheses was conducted by O'Leary, Kent, and Kanowitz, 1972.

Four undergraduate females watched specially prepared videotapes supposedly representing baseline and treatment conditions in a classroom for emotionally disturbed children. In fact, however, there were no differences in rates of disruptive behavior in the two conditions (the baseline and treatment). The study was presented to the observers as an investigation designed to replicate some earlier research on token programs in which only reinforced behaviors in the token program decreased. It was stated that other behaviors not reinforced would presumably not change.

After insuring adequate reliabilities in a pre-baseline condition, the four observers watched four baseline tapes and then four pseudo treatment tapes. One experimenter gave the four students the specific expectations regarding the outcome of the experiment, and in addition during the pseudo treatment condition, this experimenter gave the students positive or negative feedback regarding how well their data conformed to the experimental hypotheses. More specifically, he shaped their data recording by giving positive feedback to the observers only if their data conformed with the experimental hypotheses. That is, he made positive comments like, "The data really seems to be reflecting the treatment change" or negative comments like, "It's strange that you hav so many disruptive behaviors--this treatment usually works.

The observers' data were converted to difference scores between a group of four well trained observers (criterion observers) and themselves. These differences during baseline and treatment were then subject to an analysis of variance which allowed us to separate effects on the categories predicted to decrease and those predicted not to decrease. The data clearly supported the proposition that one can shape data consonant with one's experimental hypotheses if, in addition, the observers are informed of that expectation. On two categories of behavior predicted to change, the changes were reported by the observers--despite the fact that

there were no actual changes in the behavior as recorded by the criterion observers. In contrast, on those categories not predicted to change, no change was reported by the observers.

B. The Reactive Nature of Reliability Assessment

As a result of the Kass and O'Leary (1970) and Kent (1972) studies, it became apparent that observers may drift naturally in their observational criteria when different observer-groups are assigned to observe the same phenomena. Consequently, a study was conducted by Romanczyk, Kent, Diament, and O'Leary (in press) to assess whether observers would modify their recordings to match reliability checkers who adopted differing observational criteria. Throughout a study, two reliability checkers employed a unique modified version of our standard observational code. Four of the nine categories of the behavioral rating code were modified to produce stable but differential observational criteria fo the two assessors. This manipulation was intended to increase the detectability of matching by the observers of the different observational criteria employed by each assessor. As a result of these modifications, the code employed by Assessor I produced a higher frequency than the code employe by Assessor II on two categories: vocalization and noise. In employing the modified code, Assessor I would record even the softest vocalizations and also any "mouthings" the child might make as vocalizations, while Assessor II would record only the louder vocalizations and ignore such behavio as humming, whispering, and sighing. Further, the behavioral code was modified so that Assessor II would record a greater frequency than Assessor I on two other behaviors: playing and orienting. It was required that these differential observational criteria be sufficiently well-defined that the assessors would be reliable with each other at a moderate level and that this level of reliability between assessors not vary across experimental conditions. In short, an artificial difference was created between the ratings of Assessor I and Assessor II.

For two and one half weeks prior to the experiment each assessor employed his respective version of the modified code and on regular but different occassions, computed total reliability (for modified and unmodified categories combined with each observer. Reliability was computed for five observers a median of four times (range 2-4). These reliability computations provided the only opportunity for observers to note the unique observational criteria being employed by the two assessors. At no time, however, did either reliability assessor make any statement that overtly contrasted his rating criteria with those of the other assessor.

Results of this study indicated that knowledge of which reliability checking was measuring reliability produced a substantial shift in observational criteria. Thus, observers adjusted their rating criteria as a function of the feedback they received. That is, observers adopt idiosyncratic rating criteria in order to match the observational criteria of their reliability checker.

Induced expectations per se failed to influence the recordings of disruptive classroom behavior by undergraduate observers. However, the induced expectations did influence global evaluations of such change by the observers. That is, despite the fact that observers did not change their recordings to be consonant with experimenter predictions, when asked what their recordings reflected following the study, they reported that they had recorded behaviors which reflected treatment changes.

When experimental expectations were combined with shaping of data consonant with experimental hypotheses, observer recordings were markedly influenced. The implications of th data were unequivocal: one should not provide daily evaluative feedback regarding the extent to which observer's recordings are reflecting the expected treatment change.

An unanticipated problem of observer drift was incountered in this research which completely confounded the results of one expectation study (Kent & O'Leary, 1973) and which rendered a different interpretation of an earlier study by Kass (1971). The observer drift refers to a random fluctuation in the observational criteria used by groups of observers assigned to different treatment conditions. The phenomena of drift was clearly documented by O'Leary and Kent, 1973. It appears that the process of computing reliability and discussing differences in recording modifies an observer's interpretation of the behavioral code to more closely match those observers with whom he is working. When observers are divided into different groups, different modifications of the observational code may emerge. These modifications appear to have a random effect on data generated and must be differentiated from possible systematic biases due to observer expectations. The implications of the observer drift problem are very serious and O'Leary and Kent have suggested various ways of dealing with the problem.

These data suggest that it is unwise to confound individual observers or groups of observers with different experimental conditions. However, even in single group within-subject designs, there exists the possibility that observers may "drift" in their application of a behavioral code, yielding data recorded during one experimental condition incomparable to data recorded during a subsequent condition. Montrose Wolf (personal communication, 1972) has suggested a procedure of training a new group of observers several weeks after the initiation of a study and assessing their comparability to observers who have been collecting data in this setting. When no differences are found between the two

groups of observers, it is clear that drift has not occurred. However, in the absence of such information, it seems prudent to take one of several steps to avoid confounding observer drift with differential treatment interventions. In between-subject designs, one could employ a single group of observers to record data from all treatment groups. Alternately, several groups of observers could be rotated periodically from one treatment group to another. Clearly neither of these procedures guarantees that the recordings from a particular experimental condition will represent comparable applications of the behavioral code at any two points in time. This procedure does assure, however, that the data from each treatment group will be equally affected by any modifications in the behavioral code which do occur.

In within-subject designs, the critical comparisons involve one experimental condition instituted at one time and another condition instituted subsequently. Assuming that observer drift is a random phenomenon, one might employ a number of independent observer groups across all experimental conditions. For example, if experimental conditions each lasted a week or longer, different observers or different groups of observers could be employed on each day of the week. Drift among groups would thus add to the variation of data from each condition, but would not distort comparisons of one condition to another. An alternate procedure would involve videotaping the behavior of interest during all experimental conditions and showing these recordings to observers in random order. When this is impractical, observation of videotapes of a sample of behavior from each experimental condition would provide a measure of the veridicality of behavioral recording obtained <u>in vivo</u> across time.

Bibliography

Barber, T. X., & Silver, M. J. Fact, fiction, and the experimenter bias effect. Psychological Bulletin Monograph, 1968, 70, No. 6, Part II, 1-29.

Bushell, D., Wrobel, P. A., & Michaelis, M. L. Applying "group" contingencies to the classroom study behavior of pre-school children. Journal of Applied Behavior Analysis, 1968, 1, 55-63.

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

Kass, R. E. The effects of observer bias in field-experimental settings. Unpublished master's thesis, State University of New York at Stony Brook, 1971.

Kass, R. E. & 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, Lawrence, Kansas, April 9, 1970.

Kent, R. N. Expectation bias in Behavioral Observation. Unpublished doctoral dissertation, State University of New York at Stony Brook, 1972.

Madsen, C. H., Becker, W. C., & Thomas, D. R. Rules, praise and ignoring: Elements of elementary classroom control Journal of Applied Behavior Analysis, 1968, 1, 139-151.

McKenzie, H. S., Clark, M., Wolf, M. M., Kothera, R., & Benson, C. Behavior modification of children with learning disabilities using grades as allowances and tokens as back-up reinforcers. Exceptional Children, 1968, 34, 745-752.

O'Connor, R. D. Modification of social withdrawal through symbolic modeling. Journal of Applied Behavior Analysis, 1969, 2, 15-22.

O'Leary, K. D. & Becker, W. C. Behavior modification of an adjustment class: A token reinforcement program. Exceptional Children, 1967, 33, 637-642.

O'Leary, K. D., Becker, W. C., Evans, M. B., & Saudargas, R. A. A token reinforcement program in a public school A replication and systematic analysis. Journal of Applied Behavior Analysis, 1969, 2, 3-13.

O'Leary, K. D., Kent, R. N., & Kanowitz, J. Shaping data consonant with experimental hypotheses. Unpublished

manuscript, State University of New York at Stony Brook, 1972.

O'Leary, K. D., & Kent, R. N. Behavior Modification for Social Action: Research Tactics and Problems. Chapter to appear in L. Hamerlynck, et. al. (Eds.). Critical Issues in Research and Practice. Champaign, Illinois, Research Press, February, 1973.

Romancyzk, R. G., Kent, R. N., Diament, C. & O'Leary, K. D. Measuring the Reliability of Observational Data: A Reactive Process. Journal of Applied Behavior Analysis, in press.

Rosenthal, R., & Fode, K. The effects of experimenter bias on performance of the albino rat. Behavioral Science, 1963, 8, 183-189.

Rosenthal, R., & Jacobsen, Lenore. Teacher's expectancies: Determinants of pupils' IQ gains. Psychological Reports, 1966, 19, 115-118.

Scott, P. M., Burton, R. V., & Yarrow, M. R. Social reinforcement under natural conditions. Child Development. 1967, 38, 53-63.

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

Thomas, D. R., Becker, W. C., & Armstrong, M. Production and elimination of disruptive classroom behavior by systematically varying teacher's behavior. Journal of Applied Behavior Analysis, 1968, 1, 35-45.

Thorndike, R. L. Pygmalion in the classroom: A review. Teacher College Record, 1969, 70, 805-807.

Appendix

Observational Code for Disruptive Behavior

1) Out of Chair --- symbol O

Purpose: Out of chair is intended to monitor the gross motor behavior of the child removing himself from his seat entirely. When not permitted, such behavior (e.g., running around the room) may interfere with the child's learning and is potentially distracting to others.

Description: Observable movement of the child from his chair when not permitted or requested by teacher. None of the child's weight is to be supported by the chair, but the child may be in physical contact with chair.

Critical Points: None of the child's weight is to be supported by the chair.

Includes: Child is leaning on desk and has either lost all contact with the chair or none of his weight is actually being supported by the chair.

Time limits on the following beginning with teacher's permission. Allow 15 seconds for a child to get from the teacher's desk to his own. Allow 15 second for a child to return to his own seat after completing a task (i.e., placing a word card on the wall). Pencil sharpening - 1½ mins. Getting a drink - 1½ mins (fountain in room). Getting a book - 1½ mins. (time limit starts from the second that the child gets out of seat). Going to the bathroom: (A) 2 min. limit, (B) 30 sec. limit beginning when the child leaves bathroom.

Note: If the child returns to the chair after 1½ (or 2 mins., where applicable), but during the 10 sec. inter-interval period, the "O" will be recorded in the 20 sec. interval just prior to the 10 sec interval.

Going to get a reading book during a math lesson. When a child is full standing and the back of legs touch chair, or child is

fully standing and is touching back of chair with hands. Going to teacher's desk when not permitted. Throwing away papers. Stretching (if child actually leaves seat).

Excludes: Retrieval of an accidentally dropped task-related object. Leaning forward to pick up an object even if all contact with the chair is momentarily lost, providing the child is not standing fully erect on feet. Include if child begins crawling around on floor after retrieving object, also, include if child is moving from desk in a crouched position, so as not to let the teacher see him, etc.

2) Modified out of chair ---symbol Θ

Purpose: Modified out of chair is intended to monitor less intense motor behavior than displayed in out of chair, and behavior which is usually only distracting for the child himself rather than others.

Description: Movement of child from his chair, with some of his weight still being supported by the chair.

Critical Points: The child is still at his desk and some of his weight is being supported by the chair

Includes: Leaning forward to pick up an object even if all contact with the chair is momentarily lost, providing the child is not standing fully erect on feet. Bouncing in chair, e.g., in responding excitedly to some event. Kneeling on chair. Sitting on back of chair. Both feet on or in desk Lying across chair horizontally. Standing near desk with one foot on the chair.

Excludes: When child is fully standing and the back of legs touch chair. Sitting on one or both feet. One "cheek" off chair.

3) Touching other's property --- symbol T

Purpose: Touching is intended to monitor behavior which is distracting to the child and very often to others when the child comes into contact with the personal property of another.

Description: Child makes into contact with another's property without permission to do so.

Critical Points: The child does not have permission for his action and not that his action may or may not result in an alteration or post hoc permission.

Includes: Grabbing, re-arranging, destroying the property of another. Using material object as extension of hand to touch others' property. Hand brushing on others' desk if this act is incompatible with learning (i.e., the child is attending to the act). Touching desk of another, whether other person is seated in it or not (this includes teacher's desk). Resting elbows on desk behind if this act is incompatible with learning or annoys the other child.

Excludes: Touching others on the back or any part of the body or clothing. Use of shared possessions such as rulers, erasers, art materials. Elbow resting on another's desk or hand brushing against it, it the desks are together and neighbor is not disturbed and such an act is not incompatible with learning. Walking past a desk, chair, etc. and accidentally brushing or touching the desk, chair, etc., i.e., child is not attending to the behavior.

Note: When child is at teacher's desk with permission, and is waiting to be helped, do not score idle touching of objects on teacher's desk. Touching should be scored, if the teacher specifically instructs child to stop and child continues or if child is instructed to perform some task at desk and then begins to touch objects on desk.

Vocalization ---symbol V

Purpose: Vocalization is intended to monitor verbal behavior which is usually distracting to both the child and to others.

Description: For the sake of consistency, any audible non-permitted vocalization is to be recorded even though in the opinion of the observer it did not "seem" disruptive. Any

non-permitted "audible" behavior emanating from the mouth.

Critical Points: The observer must actually hear the vocalization. Inferences are not acceptable except as noted below.

Includes: If vocalization is obvious, but can't be heard (obvious - if another child responds). Answering without being called on. Moaning. Yawning. Any noise made with mouth when eating - unless child has permission to eat. Any vocalization made in response to the disruptive behavior of another child, e.g., telling another child to return stolen article, crying in response to aggression committed to his person or possessions, etc., if the child has not received permission specifically from the teacher to speak. Whispering, belching, crying, shouting, "operant" coughs or sneezes.

Excludes: Vocalization in responses to teacher's question. Sneezing. Automatic coughing.

Note: Once a child is recognized by the teacher, vocalization is not scored, regardless of content of the vocalization: crying, yelling, etc., until the teacher specifically instructs the child to stop.

Playing --- symbol P

Purpose: Playing is intended to monitor often subtle manipulative behavior that is distracting to the child and possibly also distracting to others.

Description: Child uses his hands to play with his own or community property, so that such behavior is incompatible (or would be incompatible) with learning.

Critical Points: Child uses his hands to manipulate his own or community property.

Includes: Playing with toy car when assignment is spelling. Playing with comb or pocket book. Eating only when the hands are being used - chewing gum is not rated as P unless child touches or manipulates it with his hands. Poking holes in workbook.

Cleaning nails with pencil. Drawing on self. Manipulating pencil in such a manner as to make the behavior incompatible with learning, e.g. shoving pencil back and forth on desk. Moving pencil through air as an airplane. Picking scabs, nails, or nose if the desired "object" is separated from the body and manipulated. Looking into desk and moving arms, but does not come out with a task-related object. Working with or reading non-task related material, e.g., reading page 25 when told to read page 1, doing math when told to do spelling, etc.

Excludes: Touching others' property. Playing with own clothes.

Note: Include if article is removed from body, e.g., shoes, tie, buttons, scarf, etc., and is manipulated.

Lifting desk or chair with feet (rate N if this creates audible noise). Random banging of pencil on desk (rate N if audible). Simple twiddling pencil if it is not seen as being incompatible with learning.

Note: Rate twiddling pencil, banging pencil, or putting pencil in mouth, hair, behind ear, etc., if child attends to such behavior and ceases attending to assigned task. Operational definition of attending: child either looks at manipulated object or begins to manipulate object in non-random patterns for more than 5 seconds.

Picking scabs, nails, or nose if the desired "object" is not separate from the body.

Orienting Response --- symbol ●

Purpose: Orienting is intended to monitor the gross motor behavior of turning around from the designated point of reference. Such behavior is distracting to child since it usually precludes attending to assigned task, and is often distracting to others.

Description: Child turns more than 90 degrees from poin of reference while seated.

Critical Points: The child must be in his seat; he may be in a modified position; and orienting includes both the horizontal and vertical axis.

Includes: Turning to the desk behind. Looking to the rear of the room. Turning around in chair or turning chair around. Leaning back in chair more than 90 degrees.

Note: Point of reference is typically child's desk, but may be the teacher if the children are directed to attend to her. If child should turn desk at some angle, point of reference becomes where desk was originally, not to where the child has moved it. Also, the child's chin should be used as the indicator of how far he has turned. Therefore orienting is rated when child's chin has turned more than 90 degrees from point of reference.

Excludes: Orienting during class discussions when the teacher directs (eith implicitly or explicitly) the class to attend to a child's explication of an answer. Orienting while picking up a task related object. When child is in corner or otherwise out of his chair.

Noise ---symbol N

Purpose: Noise is intended to monitor the frequency of distracting sounds produced by the child other than vocalization.

Description: Child is creating any audible noise, without permission, other than vocalization. For the sake of consistency, any audible sound is to be recorded even though in the observer's opinion it did not "seem" disruptive.

Critical Points: The observer must actually hear the sound to rate it. Inferences are not acceptable

Includes: Turning pages in an exaggerated manner, producing noise. Moving desk around. Pencil tapping. Banging of any object. Fishing in desk without coming out with anything or coming out with an inappropriate object (if noise is actually made in

the process). Shuffling feet more than once each way. Any noise made while getting out of chair without permission. In general, any noise made in conjunction with any disruptive behavior, e.g. any noise made when child throws a book or other object at another (A).

Excludes: Shuffling feet (if only once each way). Accidental dropping of a task-related object (book or pencil). Pushing chair back and forth once during a permitted act (e.g., to get a task-related object).

8) Aggression --- symbol A

Purpose: To measure the highly disruptive behavior physical assaults.

Description: Child makes an intense movement directed at another person so as to come into contact with him, either directly or by using a material object as an extension of the hand.

Critical Points: Intention is to be recorded rather than just accuracy of assault. (e.g., aggression is recorded if child throws pencil or swings at another, regardless of whether or not the pencil or motion hits the child).

Includes: Blocking others with arms or body from attaining goal (e.g., while walking up aisle). Tripping. Kicking. Throwing.

Excludes: Brushing against another (include if action is continually repeated so as to tease or annoy).

9) Time-off-task --- symbol X

Purpose: Time-off-task is intended to monitor non-attending behavior, that, if excessive, is detrimental to child's performance.

Description: Child does not do assigned work for entire 20 second interval.

Critical Points: Child makes no attending response for the entire 20 second interval. Child must only attend, i.e., "looking at," his work. Inferences that, "he isn't really thinking about it," are not acceptable.

Includes: Child does not write when so assigned. Child does not read when so assigned. Child is working on inappropriate material (e.g., on math during spelling, etc. Day-dreaming - as reflected in not working. Child does not ask teacher for additional work or help when finished with assigned task, and merely sits at desk or begins to play for entire interval. When in corner, child's head must be within a 45 degree angle from the corner formed by 2 walls (i.e., if his head is facing either of the 2 walls directly, for a 20 second period, he would be rated X).

Excludes: Child has his hand raised to ask questions. Child is told he may cease working if he so desires.

10) No inappropriate behavior as defined by the above categories --- symbol -