To the Graduate Council:

I am submitting herewith a dissertation written by Michael H. Schwartz entitled "Multi-Method Assessment of a Classroom Intervention." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Psychology.

Robert G. Wahler

Robert G. Wahler, Major Professor

We have read this dissertation
and recommend its acceptance:

M. Wyatt

John W. Edgerly

Eric Scott

Sandra Iwaniszyn

Accepted for the Council:

L. Evans Beck

Vice Chancellor
Graduate Studies and Research

Thesis
T82
S259
cop 2

MULTI-METHOD ASSESSMENT OF A CLASSROOM INTERVENTION

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Michael H. Schwartz

December 1978

1371189

ACKNOWLEDGMENTS

I would like to thank all those who helped and supported me in this project. Dr. Robert Wahler, my committee chairman, gave me continuous support. Much of the theoretical background and experimental design of the research were developed in conjunction with Dr. Eric Sundstrom. Without his guidance I could not have done this project. Drs. John Edgerly, Sandra Twardosz, and William Verplanck encouraged me and provided constructive criticism.

My wife Susan and our daughters Kate, Diana, and Alison stayed with me through it all.

ABSTRACT

A multi-method assessment system is developed and used to investigate group oriented and self-perception variables in a classroom intervention. The program developed a fifth grade teacher's monitoring and instructional skills by observing her and providing feedback. Experimental, comparison, and testing effects classes were all measured before and after the intervention. Data were analyzed via t-tests, ANOVA's, and factor analyses. Results suggest mixed effects of the intervention. The teachers' ratings were strongly influenced by the students' dominance and popularity as well as their observed behaviors. The multi-method assessment system appeared to possess a high degree of convergent validity. The data suggest the overjustification effect should be evaluated in relation to group oriented variables.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
II.	METHOD	7
	Overview	7
	Setting	8
	Subjects	10
	Treatment Program (Independent Variable)	10
	Dependent Variables	12
III.	RESULTS	19
	Data Analysis	19
	Effectiveness of Intervention	20
	Changes in Group Process as Reflected in Factor Analyses	25
	Summary of Changes from Pre to Post in Each Class	29
IV.	DISCUSSION	31
	Effect of Intervention on Class	31
	Hypotheses and Findings Relating to Individual Variables	32
	Convergent Validity of Dependent Variables	40
	Over-Justification Effect	42
REFERENCES		44
APPENDIX		48
VITA		51

LIST OF TABLES

TABLE		PAGE
1.	Overview of Experimental Design.	9
2.	Reliability of Three Dependent Measures.	15
3.	Analyses of Variance of Observational Categories Between Classes and Across Weeks	22
4.	Comparison of Pre- with Post-Scores in the Experimental and Comparison Classes and Comparison of Change Scores in the Experimental Class with Change Scores in the Comparison Class.	24
5.	Factor Loadings on First Six Factors of Factor Analyses for Experimental and Comparison Classes Before and After Intervention	26

CHAPTER I

INTRODUCTION

As understanding of classroom contingency management increases, investigators are examining more complex and subtle effects of these programs. One parameter in this development is a concern with group oriented variables. For instance, studies have examined sociometric status of students (Alden, Pettigrew, and Skiba, 1970; Drabman, Spitalnik, and Spitalnik, 1974; Acker, Oliver, Carmichael, and Ozerkevich, 1975), social interaction of students (Acker, Oliver, Carmichael, and Ozerkevich, 1975; Drabman and Lahey, 1974), the effect of reinforcement upon non-reinforced students' behavior (Christy, 1975; Kazdin, Silverman, and Settler, 1975), the use of peers as behavior modifiers (Noonan and Thibeault, 1974; Lovitt, Lovitt, and Eaton, 1973; Csapo, 1972; Solomon and Wahler, 1973), and the respective effectiveness of self, peer, or adult administered contingencies (Kreuger, 1971; Brown, 1975; Felixbrod and O'Leary, 1973).

Another parameter in the study of classroom contingency management has been the effect on students' self-perception. Lepper and Greene have developed this line of research in a series of studies (Lepper and Greene, 1975; Lepper, Greene, and Nisbett, 1973; Greene and Lepper, 1975). According to one self-perception theorist (Bem, 1972) attitudes are often inferred from a person's own overt behavior. For example, Nisbett and Valins (1972) found that men's ratings of the attractiveness of pictures of women were affected by what they thought were their heartbeats. In an

operant program, if children follow rules set by adults, the children may see themselves as rule-followers.

An exception to the hypothesis that self-perception agrees with overt behavior is the over-justification effect: A person who has a large external incentive to perform some act may see no self-motivation to do it. For example, Festinger and Carlsmith's (1959) classic study showed that subjects induced to lie for $1 later changed their attitude to agree with the lie more than subjects whose incentive was $20. It follows that behavior modification should produce changes in self-perception only when the changes in behavior are seen as self-motivated or internally controlled.

Lepper and Green (1975) demonstrated that providing an expected reward for engaging in an intrinsically interesting activity has the effect of reducing this intrinsic interest as measured behaviorally two weeks later. Also, salient observation of a child while he is engaging in the intrinsically interesting activity reduces intrinsic interest to about the same degree as an expected reward. The authors interpret this in the self-perception framework by contending that the salient reward or observation induces the individual to discount his intrinsic motivation and thus to feel externally motivated and controlled. They suggest that extrinsic rewards should be applied only when no intrinsic interest exists and then only the minimal effective amount should be used. Likewise teachers should be most chary in using observation in monitoring their students.

Lepper and Greene's line of research and theory has been strongly attacked by Reiss and Sushinsky (1976). They propose that the

"over-justification effect" or "decreased play effect" represents simply one particular manner of playing which is less enjoyable than others. That is, children might not enjoy playing as much if they are rewarded for it or if they are being watched. Reiss and Sushinsky point out that Lepper and Greene use only single trial experiments and present evidence to show that decreased play effects do not occur over repeated trials. They also cite evidence that token economies enhance rather than decrease internal attributions of control over behavior.

However, Lepper and Greene have demonstrated that whether conceptualized as an over-justification or decreased play effect a decrement in performance can be induced by external rewards or adult observation in certain circumstances.

It is the thesis of the present study that these two parameters in the study of classroom management (group oriented variables and self-perception) are intimately related and should be studied in light of each other. For example, Drabman and Lahey (1974) studied the effect of teacher provided feedback alone without any additional reinforcement, given on an individual contingency basis. They found that the subject's behavior improved as a result of feedback alone, the target was chosen sociometrically more often during feedback than baseline periods, and more positive comments were directed to the target by peers during feedback than baseline. One can speculate that the change in peer behaviors toward the subject altered his self-perception.

These findings indicate a need to delineate the effect of various monitoring and reinforcement techniques upon both group process and self-perception. A methodology for accomplishing this has been developed

and utilized in this study. It concurrently assesses self-perception and group oriented variables. By simultaneously measuring both types of variables one can hopefully judge each in light of the other. The logic of this approach is that of convergent validity which is "represented in the agreement between two attempts to measure the same trait through maximally different methods" (Campbell and Fiske, 1959). Thus a student's self-report of rule following is compared to the teacher's rating of him and to observations of him by trained observers.

In their treatment of validity in relation to naturalistic behavioral data, Johnson and Bolstad (1973) concur with Campbell and Fiske (1959). In addition, they raise the problem of the generalizability of data across occasions in the same setting and across settings. In order to assess the generalizability of the data generated in the present study, as well as to control for threats to validity, we collected data before and after a classroom management intervention in both the intervention classroom and a comparison classroom. I then factor analyzed each of the four resulting data sets. This procedure related the variables to each other in different settings and at different times, thus assessing both their convergent validity and generalizability across occasions and settings.

The present study employs maximally different methods to measure relevant group oriented and self-perception variables. The contingency management intervention under investigation strives to increase work, appropriate social interaction, and compliance to classroom rules. The dependent variables reflect this. In order to achieve maximally different methods, I had outside observers, the students, and the teachers

provide data. The observers recorded social, work, and rule relevant behaviors of the students. The students rated each other's dominance and popularity, their own behavior and attitudes regarding rule following, and they told stories about rule relevant incidents. In using the stories, I hoped to assess the students' view of their school environment with minimum reactivity. The teachers rated the students on five dimensions: responsible, independent, popular, aggressive, dominant. I hoped to evaluate each variable in terms of the others. Thus students' self-report of rule following and taking responsibility can be compared with the teacher's rating of them and with actual observations of their behavior.

The first general expectation of this study was that the intervention would be effective in terms of the traditional measures. That is, work would increase and rule breaking would decline. A second expectation was based upon Patterson and Reid's (1970) distinction between "coercion," the prepotency of negative reinforcement contingencies in a social system, and "reciprocity," a social system based upon the reciprocal exchange of positive reinforcers. In the social psychology literature there are some very similar distinctions: reward and coercive power versus expert power (French and Raven, 1959), normative versus informational influence (Deutsch and Gerard, 1955). See also Kelley and Thibaut's (1968) discussion of group process.

I expected an effective intervention to change the classroom social system away from coercion toward reciprocity. This might be reflected in the dependent variables in several ways: Aggressive behavior would decline and appropriate interaction increase.

The students would include less aggression and rule breaking in their stories. Students would report taking more responsibility for their behavior. The level of popularity, shown by the number of friendship choices, would increase. The teacher would rate the students in a more positive direction.

Third, I expected the measurement system to display convergent validity. That is, the variables would be easily interpretable in relation to each other in the factor structure and be congruent. By "congruent" I mean the most predictable relation between the variables. Thus the teachers' rating, responsible, to be congruent would relate positively with both the students' observed work and the students' report of their own behavior and attitudes and relate negatively with aggressive behavior. Another example of congruence would be that students rated dominant by their peers would be rated dominant by the teacher and also give more orders to their peers.

Fourth, I expected the variables in each of the four data sets to relate to each other similarly in their respective factor structures. This would increase one's confidence in the convergent validity of each variable. However, I also hoped the factor structures would reflect changes in group process introduced by the classroom management intervention.

Finally, I expected to generate new hypotheses about some of the variables. In particular, I hoped comparison of group oriented and self-perception variables would be relevant to the over-justification effect controversy.

CHAPTER II

METHOD

I. OVERVIEW

We measured a fifth grade elementary school class with 25 students before, during, and after a behavior modification intervention. I designated another fifth grade class with 25 students as a comparison group, and we measured it at the same times. Observation and interview data were obtained. We used three kinds of instruments in the interviews: sociometric measures, teachers' ratings, and self-report measures. I used two sociometric measures: dominance hierarchies and choice of friend matrices. Teachers' ratings of the children on five dimensions measured their judgments of the children. I used the following three self-report measures: first, children described their rule following behavior in a sentence completion format; second, in a forced choice format each child decided whether hypothetical negative outcomes were "my fault" or "his fault" and whether he followed class rules in hypothetical situations because he "wanted to" or because he "had to"; third, a projective test elicited the children's stories about relevant aspects of their social environment.

Observations for both experimental and comparison classrooms were conducted every day for a half hour during an academic period. Baseline consisted of eight days of observation. During this period we interviewed the students and teachers. Then, the intervention began. On the

fourth week of the intervention we interviewed the students and teachers again. In regard to the interviews, because random assignment of subjects was impossible, this is a quasi-experimental design and meets the definition of a "non-equivalent control group design" (Campbell and Stanley, 1963). Another feature of the design controlled for pretesting effects in the self-report measures; a third, "pretesting effects" classroom was used for this purpose. It was another fifth grade class at the same school with 27 students. Half (14) of these students were randomly selected. When we interviewed the students of the experimental and comparison classes, the self-report measures were given to this random sample in the pretesting effects class. Then on the fourth week after the intervention began we interviewed all 27 students of the pretesting effects class. The scores for the non-pretested students were compared to those of the pretested students to determine if there were any effects of pretesting. Table 1 outlines the method.

II. SETTING

The setting was a fifth grade class of an elementary school (grades K-6) in a predominantly black area of Knoxville, Tennessee. It was a high risk school in frequency of child delinquency cases and other Juvenile Court referrals. We selected the experimental class for intervention because the principal of the school designated it as the class with the most "problem" children. The class was located with the other two fifth grade classes in an older building (circa 1900) about fifty feet behind a larger structure which houses most of

TABLE 1

OVERVIEW OF EXPERIMENTAL DESIGN

Number of Subjects Used in Each Measurement Category				
	Pre-test		Post-test	
Type of Measurement	Experimental & Comparison Classes	Pre-testing Effects Class	Experimental & Comparison Classes	Pre-testing Effects Class
Observations	25		25	
Sociometric Measures	25		25	
Teachers' Ratings	25		25	
Students' Self-Reports	25	14	25	27

Schedule in Weeks of Observations and Interviews						
	Pre-test Interviews				Post-test Interviews	
1	2	3	4	5	6	7
Baseline	Treatment					Last Week of School Year
Weeks of Observation						

the school. It was on the second floor and overlooked the school's asphalt play area behind. The pretesting effects classroom was adjacent to the experimental classroom. The comparison classroom was on the first floor of the same building.

III. SUBJECTS

Experimental subjects were the 25 students in the experimental class. Comparison subjects were the 25 students in another fifth grade class. The third "pretesting effects" subjects were the 27 students in the other fifth grade class. The experimental class had 18 boys and 7 girls, ages 10-12 (with one 15 year old). The comparison class contained 10 boys and 15 girls, ages 10-12. The pretesting effects class contained 13 boys and 14 girls, ages 10-12. All three teachers were young females and the principal a somewhat older male. All three classes also had a student teacher in the morning. The principal, teachers, student teachers and all but five of the students (distributed about equally among the classes) were black.

IV. TREATMENT PROGRAM (INDEPENDENT VARIABLE)

We intended to demonstrate the efficiency of developing the teacher's skills in managing the class. A point system, time out procedures, and other specific aspects of classroom management were thus subordinate to the primary goal of improving the teacher's skills. Using this approach, we hoped to minimize consulting time and use no money for back-up reinforcers.

After a two-week baseline observational period, intervention

began and continued for four weeks until May 30. On May 5, we discussed the class rules with the teacher. Together we revised the rules to still prohibit serious infractions but give more leeway to the students in certain areas such as talking quietly to a neighbor. Secondly, we brought her desk into the center of the room and placed four problem children's desks directly in front of hers. Thirdly, we discussed the teacher's monitoring and command style with her. Together we worked out guidelines for her to use. We made copies of the guidelines in classroom management for the teacher and intervention staff. The guidelines prompted the teacher to record each assignment on the blackboard. When giving instructions she was to stand facing the class, clearly specify the rule or command, speak clearly enough to be heard at the back of the class, and give the instruction so that anyone could easily understand what was required. She was to respond non-critically to appropriate approach behavior. In addition she was to intermittently walk around the class when children were doing a work assignment and give social approval to on-task students. For those who were finished, she was to make sure they had something to occupy themselves.

For the first ten days of the intervention, two undergraduate assistants observed the teacher for an hour a day. For the second ten days each undergraduate observed for one-half hour a day. They provided the teacher with written feedback emphasizing about 50% good points and 50% bad points.

On May 7, time out procedures were instituted. We felt that the teacher carried them out consistently and effectively. On occasions

when time out was not effective, students were sent to the principal who often paddled them.

On May 9, work record cards were instituted. The teacher taped one to each desk each morning. The student then recorded all the work he completed during the day. The teacher randomly monitored this process to prevent cheating. At the end of the day the teacher recorded the amount of work for each student. Students then received extra recess or free time in the classroom as a reward.

During the sixth week of observation (the fourth week of intervention), the experimental class teacher was ill for two days and was replaced by a substitute teacher. We obtained our post test data this week.

Our staff anecdotally reported that the teacher gained confidence and enthusiasm and by the end of intervention demonstrated greatly improved classroom management skills. They felt, however, that she was still somewhat deficient in positively reinforcing desirable behaviors.

At the end of the sixth week of observation, the undergraduate assistants left. The teacher continued the intervention program during the next week on her own, and we observed the students. This was the last week of the school year.

V. DEPENDENT VARIABLES

Observations of Students

Category system. The four observers used a thirteen category system derived from Wahler's category system (Wahler, House, and

Stambaugh, 1975). I modified Wahler's system in several ways in order to focus on the group as well as on individuals. First, I reduced the number of categories. The resulting category system emphasizes social and deportment-relevant behaviors. Secondly, we used "scan sampling" rather than observation of one subject. That is, the observer recorded the ongoing activity of each individual in a group consecutively. When the behavior was recorded, the observer also wrote the initials of the subject observed. We sequentially observed each adjacent subject until the entire group had been observed and then repeated the cycle or "scan." Each observation consisted of a five second observe, then a five second record period.

Thirdly, only one behavior was recorded for each interval as opposed to scoring several behaviors and/or stimulus categories during one interval. For this study, the first behavior which was clearly scorable during the five second interval was scored at the end of the interval. Fourthly, categories were mutually exclusive.

The resulting system measures a predominant ongoing activity or "state" rather than a variety of more molecular behaviors or "events." Thus we know the percentage of time the group spent in each category. See the Appendix for definitions of observation categories.

Observer training. Observations were conducted by four undergraduate field assistants. Approximately 15 hours of training were needed for all four observers to reach criterion level of reliability (80%). Training began with the observers reading the definitions found in the Appendix. Next, this category system was used in scoring a

series of programmed videotapes depicting child behavior at different levels of complexity. Then observations began in the experimental and comparison classrooms. After one week of observing for one-half hour every day, all four observers had reached criterion level on all categories. Weekly retraining sessions using the videotapes and with the experimenter as an additional reliability check were conducted throughout the study.

Observer reliability. Reliability checks were made weekly in both classrooms. Unfortunately, because of scheduling and transportation difficulties, random assignment of combinations of observers for reliability checks was impossible. Thus, the same two observers always served as each other's reliability check. I attempted to minimize the danger of "observer drift" through the weekly retraining sessions and reference to the written definitions.

Reliability was computed in terms of percentage agreement by dividing the number of agreements on occurrence by the number of agreements on occurrence plus disagreement on occurrence. Reliability figures are presented in Table 2.

Preliminary analysis of observational data. Following each observation, the total number of occurrences of each behavior of each subject was tabulated. At the end of each week the total of each behavior for each individual and for the class was then calculated. These figures for each class were divided by the total number of observational intervals for the week for the class to yield weekly percentage occurrence figures for each behavior in that classroom.

TABLE 2

RELIABILITY OF THREE DEPENDENT MEASURES

Observational Categories			
Category	Effective Percent Agreement	Category	Effective Percent Agreement
Solitary Non-Task	.85	Non-Physical Aggression	.50
Work	.87		
Mand Adult	***[a]	Physical Aggression	.75
Mand Child	1.00		
Compliance Adult	.81	Receiving Aggression	***[a]
Compliance Peer	.80		
Interaction Adult	.76	Rule Breaking	.91
Interaction Child	.92	Punishment	1.00

Self-Description of Rule Following Behavior[b]			
Question	Effective Percent Agreement	Question	Effective Percent Agreement
1	.87	6	.97
2	.87	7	.86
3	1.00	8	.87
4	.92	9	.90
5	.85	10	.92

Description of Social Environment[b]			
Category	Effective Percent Agreement	Category	Effective Percent Agreement
Aggression	.91	Punishment	.63
Rule Breaking	.80	Authority	.85

[a]Did not occur during reliability checks.

[b]Pairwise agreement among three raters.

Further calculations were then based upon these weekly percentage occurrence figures.

Sociometric Measures

Dominance hierarchies. We obtained from each student a dominance hierarchy of his class. First we assured each student that what he said would remain confidential. Then we handed him a list of the students in his class and asked him to choose the students most and least "able to get his own way." We wrote the names of the most and least dominant students near the top and bottom of a blank sheet of paper and placed it so that the subject could easily reach it. We then asked him to rank the others between the two extremes, and their names were read one at a time in a random order. After we read each name, we placed it as the subject directed between the high and low extremes. After we read the last name, we asked the subject to look at the hierarchy he constructed and make sure it was correct. If not, we made the indicated changes. We then averaged each subject's rank on every list to yield a dominance rank score.

Choice of friend matrices. We read to each subject in random order the names of the other children in the class. He indicated those who were his friends. A sociometric matrix resulted. From this matrix can be seen the total number of choices between children. Thus each subject has a score for the times he is chosen, the number he chooses, and the number of children who reciprocate his choosing them.

Teachers' Ratings

We obtained the teachers' judgments as ratings on five seven-point rating scales. Each scale represents a different dimension with a rating of one meaning that the student has very little of that quality and seven representing the presence of a great deal of it. The five dimensions were aggressive, dominant, popular, responsible, and independent.

Self-Report Measures

Self-description of rule following. We presented five situations to each child and asked how he would respond. For instance, "If _____ didn't like what another kid said to him he would _____." We substituted the child's name for the first blank. Each of three interviewers then judged responses to be rule following, neutral, or oppositional based upon his knowledge of the rule systems of the child's class. Rule following responses were scored +1, neutral responses 0, and oppositional responses -1. Reliability of the interviewers' classification of responses is reported in Table 2.

Wanted to-Had to/My fault-His fault. The "Wanted to-Had to" measure uses five questions as to whether the child "wanted to" or "had to" follow a rule. For example, "You knock a stack of papers off a desk so you pick them all up and put them back like you found them. Did you want to or did you have to?" "Wanted to" responses were scored 1 point, "had to" responses were scored 0 point. Thus a high score indicates the child reports following rules voluntarily.

The "My fault-His fault" measure presents each subject with five

situations in which a rule is violated. We then ask him whether it is his fault or someone else's fault. For example, "You start teasing a kid who has a funny haircut. Everyone joins in and pretty soon he starts crying. Whose fault is it, yours or theirs?" "My fault" responses were scored 1 point, "his fault" or "their fault" responses were scored 0 point. A high score indicates the child accepts responsibility for rule violations.

Description of social environment. We presented five pictures to each student. The pictures showed a rule breaking or aggressive incident at the school. They were somewhat ambiguous. A particular scene might be a rule violation or an innocent occurrence. Another could be an accident or intentional aggression. Prior to use in this study, five students and five teachers inspected the cards. They all said the drawings realistically depicted common scenes at the school.

We told the student to tell what led up to the picture, what was happening right there, and how it would end. The interviewers then scored the resulting stories as to frequency of aggression, rule brakking, authority, and punishment. Reliability of the interviewers' scoring is reported in Table 2.

For brevity during the rest of this thesis, description of social environment will often be referred to as DOSE.

CHAPTER III

RESULTS

I. DATA ANALYSIS

Each observational dependent variable was analyzed via a two-factor (class and weeks) mixed design ANOVA with repeated measures on one factor (weeks) (Huck, Cormier, and Bounds, 1974, p. 106). In addition, the observational data of the classes were compared via a multivariate analysis of variance for groups.

Two-tailed univariate t-tests and multivariate t-tests were used to compare pre with post scores of all variables in each class and to compare ch nge scores in each class with the corresponding change score in the other class. Because the number of dependent variables precluded their being analyzed together, a multivariate t-test was used with the interview variables and another with the observational variables. With the pretesting effects data, the post scores of the pretested subjects were compared with the scores of the post-test-only subjects via univariate two-tailed t-tests.

Factor analyses were computed on the experimental class pre-intervention data set, the experimental class post-intervention data set, the comparison class pre-intervention data set, and the comparison class post-intervention data set. The interview data and the observational data for the respective week were combined in each analysis. A principal component analysis with varimax rotation was done by BMDP 4M computer program. Ten factors with eigenvalues over one were extracted

from three of the data sets while the pre-experimental data set yielded nine. The first six factors in each analysis were interpreted.

II. EFFECTIVENESS OF INTERVENTION

Although the MANOVA indicates a significant difference between the experimental and comparison class data (Hotelling-Lawley's Trace: df = 13/336, F = 10.36 $p < .0001$. Pillai's Trace: df = 13/336, F = 10.36 $p < .001$), the ANOVA's (see Figure 1 and Table 3) indicate very mixed results of the classroom intervention program. ANOVA's 1-5 (Work, Interaction Peer, Mand Adult, Mand Peer, Receiving Aggression) indicate a significant effect across time, but there is no significant difference between the classes. The trends of Work and Interaction Peer reflect much less work and more social interaction during the last week of school. ANOVA's 6-10 (Solitary Non-Task, Compliance Peer, Compliance Adult, Non-Physical Aggression, Rule Breaking) all indicate significant effects between groups, across time, and as an interaction between these two factors.

Solitary Non-Task behavior is more prevalent in the comparison class and decreases less there with the passage of time. Compliance Peer is greater in the experimental class and increases more there. The trends in both of these observational categories are equivocal in regard to the effectiveness of the intervention.

There is more compliance to commands of the teacher (Compliance Adult) in the experimental class than in the comparison class, and it increases more during the intervention than in the comparison class. In addition, Compliance Adult increases most in the last week of school,

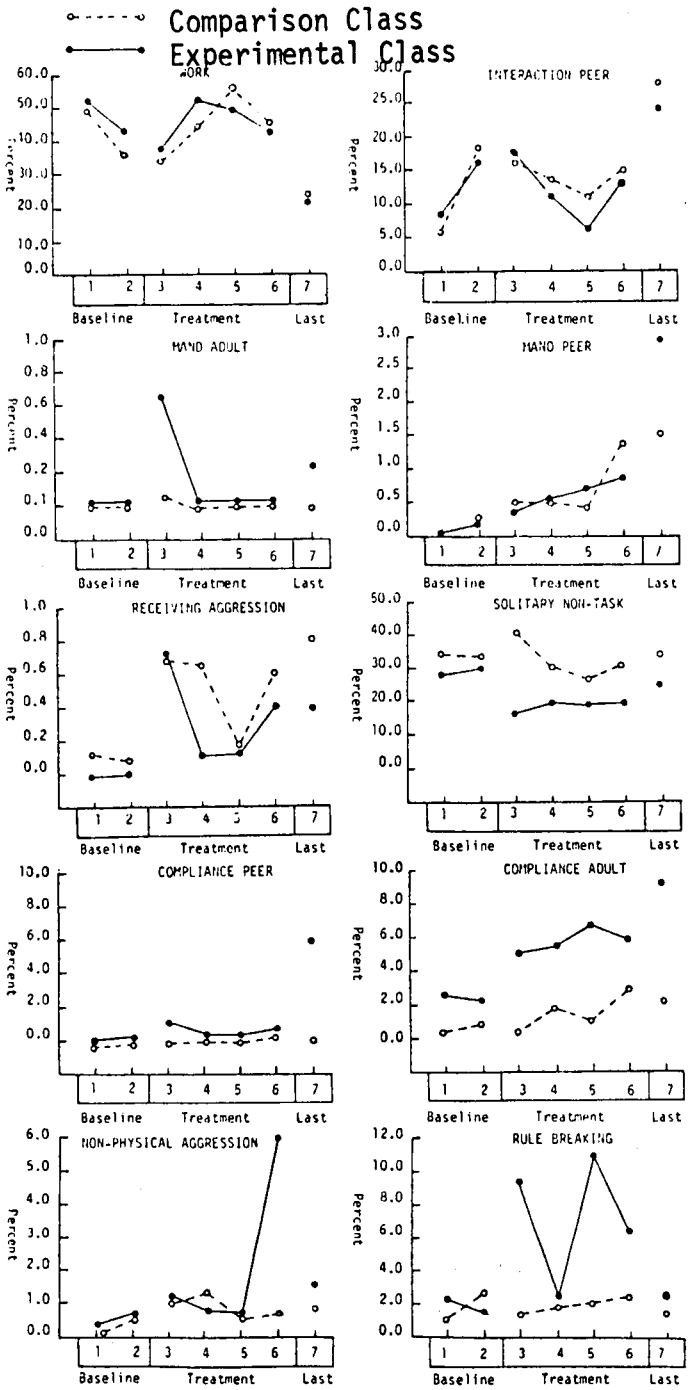


Figure 1. Percentage of total observed intervals recorded in individual categories of behavior for seven consecutive weeks.

Note. The categories presented in this figure yielded significant results in their ANOVA's. Three categories are not included. The results for interaction Adult and Physical Aggression were not significant. Punishment could not be computed because of its low frequency.

[a]Last = last week of school year.

TABLE 3

ANALYSES OF VARIANCE OF OBSERVATIONAL CATEGORIES
BETWEEN CLASSES AND ACROSS WEEKS

Source	df	MS	F
Work			
Between Subjects			
Class	1	176.57	0.3
Error Between	48	658.99	
Within Subjects			
Weeks	6	4960.38	25.2****
Interaction	6	324.80	1.6
Error Within	288	196.78	
Mand Adult			
Between Subjects			
Class	1	1.21	2.7
Error Between	48	.45	
Within Subjects			
Weeks	6	.93	2.4*
Interaction	6	.63	1.7
Error Within	288	.38	
Receiving Aggression			
Between Subjects			
Class	1	39.1	2.2
Error Between	48	17.7	
Within Subjects			
Weeks	6	38.3	2.9**
Interaction	6	.59	0.4
Error Within	288	1.30	
Compliance Peer			
Between Subjects			
Class	1	94.74	16.1***
Error Between	48	5.88	
Within Subjects			
Weeks	6	61.35	10.8****
Interaction	6	63.52	11.2****
Error Within	288	5.67	
Non-Physical Aggression			
Between Subjects			
Class	1	73.14	8.1**
Error Between	48	8.98	
Within Subjects			
Weeks	6	49.82	7.8****
Interaction	6	50.08	7.8****
Error Within	288	6.42	
Interaction Peer			
Between Subjects			
Class	1	164.84	0.8
Error Between	48	218.06	
Within Subjects			
Weeks	6	2046.25	20.9****
Interaction	6	105.90	1.1
Error Within	288	97.78	
Mand Peer			
Between Subjects			
Class	1	2.55	0.5
Error Between	48	5.38	
Within Subjects			
Weeks	6	27.93	6.8****
Interaction	6	4.97	1.2
Error Within	288	4.11	
Solitary Non-Task			
Between Subjects			
Class	1	8698.08	26.4****
Error Between	48	328.78	
Within Subjects			
Weeks	6	550.63	3.5**
Interaction	6	412.89	2.6*
Error Within	288	158.13	
Compliance Adult			
Between Subjects			
Class	1	1260.46	30.9****
Error Between	48	40.75	
Within Subjects			
Weeks	6	115.09	4.3***
Interaction	6	55.33	2.1*[a]
Error Within	288	26.61	
Rule Breaking			
Between Subjects			
Class	1	995.58	20.9****
Error Between	48	47.61	
Within Subjects			
Weeks	6	187.68	7.5****
Interaction	6	183.74	7.4****
Error Within	288	24.98	

[a] $p < .0550$ a marginal significance; * $p < .05$; ** $p < .01$; *** $p < .001$; **** $p < .0001$.

Note. All categories presented in this table yielded significant results. Three categories are not included. The results for Interaction Adult and Physical Aggression were not significant. Punishment could not be computed because of its low frequency.

after the intervention team had left. These data suggest the intervention has increased the teacher's ability to control the behavior of her students.

Non-Physical Aggression is greater in the experimental class and increases more across time. Much of the increase is during Week 6 when there was a substitute teacher for two of the days. Rule Breaking is greater in the experimental group and increases more there. It increases sharply after the intervention begins. The trend however is erratic. During the last week of school, Rule Breaking has returned to approximately baseline level. These data, in contrast to Compliance Adult, suggest the intervention has been ineffective.

Comparison of the experimental class observational data of Week 2 with Week 6 via a multivariate t-test indicates a significant change (Hotelling $T^2 = 201.02$, $p < .0001$). Although the observational data in the comparison class also changed (Hotelling $T^2 = 60.01$, $p < .05$), there was a significant difference in the amount of change (Hotelling $T^2 = 57.44$, $p < .001$). However univariate t-tests of individual variables, as shown in Table 4, indicate that the direction of the changes was often not in the predicted direction. Conclusions from these comparisons are similar to those to be drawn from the ANOVA's. In addition, they indicate the experimental class teacher has increased Punishment of her children, and this increase is greater than the increase in the comparison class. In contrast, Punishment declines in the comparison class.

Comparing the interview data from Week 2 of baseline with treatment Week 4 via multivariate t-tests indicates that as a whole neither the experimental nor the comparison class changed (experimental class

TABLE 4

COMPARISON OF PRE- WITH POST-SCORES IN THE EXPERIMENTAL AND COMPARISON CLASSES AND COMPARISON OF CHANGE SCORES IN THE EXPERIMENTAL CLASS WITH CHANGE SCORES IN THE COMPARISON CLASS

	Pre vs. Post Experimental Class			Pre vs. Post Comparison Class			Experimental vs. Comparison Class Difference Scores		
Dependent Measures	Mean Pre	Mean Post	t[a]	Mean Pre	Mean Post	t[a]	Mean[b] Dif-Exp	Mean[c] Dif-Comp	t[d]
Sociometric									
Dominance	13.2	13.1	-0.28	15.5	12.5	-0.89	-0.1	-3.0	0.86
Chosen as Friend	19.3	19.2	-0.21	14.8	15.0	0.34	-0.1	0.2	-0.40
Friendship Choices	19.3	19.0	-0.26	14.8	15.0	0.31	-0.2	0.2	-0.39
Mutual Choices	15.9	15.3	-0.82	10.8	10.2	-1.04	-0.6	-0.6	-0.04
Teachers' Ratings									
Aggressive	5.0	5.2	0.72	3.2	3.2	0.00	0.1	0.0	0.41
Dominant	4.4	4.9	2.01	2.8	3.1	1.62	0.5	0.3	0.52
Popular	4.7	5.0	1.57	4.4	4.2	-1.55	0.3	-0.2	2.18*
Responsible	5.1	5.4	2.22*	3.6	3.6	0.49	0.4	0.1	1.22
Independent	4.8	5.4	3.22**	3.8	3.7	-0.46	0.6	-0.1	2.61*
Students' Self Reports									
My fault/His fault	3.5	3.1	-1.19	3.4	3.1	-1.13	-0.4	-0.3	-0.29
Wanted to/Had to	2.9	3.2	0.85	2.2	2.6	1.41	0.3	0.4	-0.17
Self Description of Rule Following	4.4	4.4	0.00	4.0	4.3	1.00	0.0	0.32	-0.44
DOSE[e] Aggression	2.9	2.5	-1.04	3.2	2.4	-3.08**	-0.4	-0.76	0.94
DOSE[e] Rule Breaking	1.9	1.4	-2.18*	1.4	1.7	1.98	-0.5	0.36	-2.93**
DOSE[e] Punishment	0.2	0.5	1.14	0.6	0.4	-0.96	0.2	-0.2	1.49
DOSE[e] Authority	1.3	2.0	2.52*	2.1	2.4	1.28	0.7	0.3	1.05
Observational Categories									
Solitary Non-Task	30.3	20.9	-3.01**	34.8	31.6	-1.04	-9.4	-3.2	-1.40
Work	43.8	40.8	-0.75	37.7	42.0	1.68	-3.1	4.3	-1.53
Mand Adult	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mand Child	0.2	0.9	1.74	0.2	1.4	2.37*	0.7	1.1	-0.78
Compliance Adult	2.7	5.8	2.73*	1.3	3.1	2.52*	3.2	1.8	1.01
Compliance Peer	0.0	0.5	1.35	0.0	0.4	1.72	0.5	0.4	0.12
Interaction Adult	3.8	4.2	0.42	3.7	3.8	0.13	0.4	0.1	0.30
Interaction Peer	16.6	11.5	-1.98	18.2	13.5	-2.69*	-5.0	-4.8	-0.08
Physical Aggression	0.5	5.9	4.77***	0.5	0.5	0.29	5.4	0.1	4.53***
Non-Physical Aggression	0.3	0.7	1.05	0.4	0.6	0.64	0.4	0.2	0.43
Receiving Aggression	0.0	0.4	1.81	0.1	0.7	1.90	0.4	0.6	-0.45
Rule Breaking	1.6	7.0	4.27***	2.7	2.4	-0.22	5.3	-0.2	3.36**
Punishment	0.1	3.2	2.60*	0.4	0.0	-1.27	3.1	-0.4	2.82**

[a]Two-tailed correlated t-tests were used.

[b]Mean difference score, experimental class.

[c]Mean difference score, comparison class.

[d]Two-tailed independent samples t-tests were used.

[e]DOSE = description of social environment.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Hotelling T^2 = 58.64, p = n.s., comparison class Hotelling T^2 = 65.39, p = n.s.). In addition, the change in each class did not differ from each other (Hotelling T^2 = 29.10, p = n.s.).

However, as indicated by the univariate t-tests shown in Table 4, some of the teachers' ratings and the students' description of social environment did change. The experimental class teacher rated her students as more Responsible and Independent during the post test than during the pretest. And the change of the rating, Independent, was greater in the experimental than in the comparison class. Similarly the experimental teacher's rating of her students' popularity (Popular) changed more positively than the comparison class teacher's. Thus the intervention appears to have induced the experimental class teacher to rate her students more positively on several dimensions.

III. CHANGES IN GROUP PROCESS AS REFLECTED IN FACTOR ANALYSES

In the pre-experimental data set as shown in Table 5, factor one reflects responsibility. Although the teacher's rating, Independent, is the most highly correlated with this factor, the observational variables associated with it pertain to responsibility instead. In the second factor, the teacher's ratings most closely relate to the sociometric measure, Dominance. It also suggests that in the experimental class before the intervention the dominant children did not voluntarily follow the rules (Wanted to). Factor three indicates that before the intervention the teacher punished most frequently children who gave orders to other children (Mand Peer). Factor four suggests that children who described themselves as rule following (Self-Description

TABLE 5

FACTOR LOADINGS ON FIRST SIX FACTORS OF FACTOR ANALYSES FOR EXPERIMENTAL AND COMPARISON CLASSES BEFORE AND AFTER INTERVENTION

Factor	Setting: Pre Experimental Class	Post Experimental Class	Pre Comparison Class	Post Comparison Class
1	Responsible (18%)	High Teacher Ratings (22%)	Punished Popular (14%)	Socially Active (17%)
	TR[a] Independent .88	TR Popular .90	TR Dominant .91	TR Independent .93
	TR Responsible .82	TR Dominant .86	TR Aggressive .85	TR Responsible .90
	O[b] Solitary Non-Task Behavior -.76	Dominance .84	O Punishment .72	Dominance .78
	O Non-Physical Aggression -.67	TR Aggressive .81	TR Popular .72	Chosen as Friend .70
	O Work .65	TR Responsible .76	Chosen as Friend .49	TR Popular .63
	O Rule Breaking -.64	TR Independent .70		Mutual Choices .60
	Dominance .48	Chosen as Friend .51		
		DOSE[c] Authority .44		
2	Dominant (17%)	Punished (11%)	Responsible Popular (12%)	Oppositional (16%)
	TR Dominant .91	O Work -.81	TR Responsible .89	O Rule Breaking .89
	TR Popular .83	O Punishment .79	TR Independent .88	O Physical Aggression .86
	TR Aggressive .81	O Interaction Adult .62	TR Popular .55	TR Aggressive .76
	Dominance .69		My fault .49	TR Dominant .73
	Wanted to -.48		Chosen as Friend .43	O Compliance Adult .70
				TR Popular .57
3	Punished (11%)	Popular (10%)	Chooses Others (11%)	Responsible (10%)
	O Punishment .90	Mutual Choices .81	Friendship Choices .94	O Work .89
	O Mand Peer .82	Friendship Choices .78	Mutual Choices .80	O Solitary Non-Task Behavior -.75
		O Compliance Peer -.73	O Mand Peer -.52	O Non-Physical Aggression -.49
		Chosen as Friend .50	O Work .42	DOSE Rule Breaking .44
4	Self Described Rule Followers (11%)	Describes Aggression (8%)	Isolate (11%)	Dominant (9%)
	Self Description of Rule Following .83	DOSE Aggression .91	O Interaction Peer -.86	O Mand Peer .89
	DOSE Rule Breaking .77	Wanted to -.82	O Solitary Non-Task Behavior .83	O Interaction Adult .78
	Friendship Choices -.63		Self Description of Rule Following .40	TR Dominant .40
5	Popular (10%)	Verbally Aggressive (8%)	Oppositional (10%)	Scapegoat (9%)
	Mutual Choices .83	O Mand Peer .88	O Non-Physical Aggression .77	Friendship Choices .84
	Chosen as Friend .82	O Non-Physical Aggression .75	O Work -.71	DOSE Rule Breaking -.70
	O Solitary Non-Task Behavior -.45		O Physical Aggression .70	O Receiving Aggression .59
			O Rule Breaking .57	Mutual Choices .51
6	Describes Authority (9%)	Responsible (8%)	Describes Punishment (10%)	Submissive (8%)
	DOSE Authority .87	My fault .85	DOSE Punishment .80	Self Description of Rule Following .79
	DOSE Punishment .73	O Solitary Non-Task Behavior .75	Wanted to .78	DOSE Punishment .63
	O Interaction Adult -.57		DOSE Authority .54	O Compliance Peer .61
				O Non-Physical Aggression -.51
	Σ[d] = 76%	Σ = 67%	Σ = 68%	Σ = 69%

[a]TR = Teachers' Ratings.

[b]O = Observed.

[c]DOSE = Description of Social Environment.

[d]Σ = Percent of variance explained by first six factors of that factor analysis.

of Rule Following) described a lot of rule breaking at the school (DOSE Rule Breaking) and chose few other students as friends (Friendship Choices). Factor five indicates the popular children (Mutual Choices and Chosen as Friend) engaged in less Solitary Non-Task behavior. Factor six indicates that children who reported authority and punishment in their school environment (DOSE Authority and DOSE Punishment) interacted less with the adult (Interaction Adult).

In the post-experimental data set factor one suggests a strong halo effect in the teacher's ratings. She does not discriminate among the individual rating dimensions. Instead, it appears she rates the dominant (Dominance) and popular (Chosen as Friend) children highest on all dimensions. However, she appears to be more accurate in her ratings of these two dimensions than she was during pretesting. It is also interesting to note that the children she rates highest tend to report more authority in their social environment (DOSE Authority). Factor two indicates that after four weeks of the intervention the experimental class teacher punishes (Punishment) and interacts most (Interaction Adult) with the children who work the least (Work). This suggests that she is punishing appropriately, but unfortunately it also suggests she is interacting most with the children she punishes. Thus she is probably deficient in interacting and giving social reinforcement to the children who work more. Factor three suggests that the popular children (Mutual Choices, Chosen as Friend) either comply less with commands of their peers (Compliance Peer) or receive fewer commands. Factor four indicates that the students who report most aggression in their environment (DOSE Aggression) take less responsibility for

following the rules (Wanted to). Factor five shows that students who command their peers (Mand Peer) are also the most verbally aggressive (Non-Physical Aggression) toward them. Factor six shows that children who take responsibility for negative occurrences (My fault) engage in more solitary non-task behavior.

In the pre-comparison data set, factor one shows that the teacher punishes (Punishment) most the children she rates as most dominant (TR Dominant) and aggressive (TR Aggressive). These children are the most popular (Chosen as Friend) and the teacher rates them as such (TR Popular). Factor two indicates that the teacher rates as Popular, Independent, and Responsible those students who are popular (Chosen as Friend) and indicate they take responsibility for negative events (My fault). Factor three suggests that the students who chose others the most (Friendship Choices) gave fewer orders to other children (Mand Peer) and tended to do more school work (Work). Factor four indicates that students who interacted in the classroom with their peers the least (Interaction Peer) engaged most in Solitary Non-Task behavior and tended to describe themselves as rule following (Self-Description of Rule Following). Factor five reflects oppositional, anti-social behavior (Non-Physical Aggression, Physical Aggression, Rule Breaking). Factor six indicates that students who report punishment in their social environment (DOSE Punishment) also want to follow the class rules (Wanted to) and describe authority in their environment as well (DOSE Authority).

In the post-comparison data set, factor one shows the teacher has rated the dominant (Dominance) and popular (Chosen as Friend)

students as the most Independent, Responsible, and Popular. In factor two she rates as Aggressive, Dominant, and Popular the students who break rules (Rule Breaking) and are physically aggressive (Physical Aggression) to their peers. These students also comply to her requests the most (Compliance Adult). Factor three shows students who work (Work) most engage in Solitary Non-Task behaviors least. They also tend to report more rule breaking in their environment (DOSE Rule Breaking) and are less verbally abusive to their peers (Non-Physical Aggression). Factor four indicates that students who give orders to their peers the most (Mand Peer) also interact with adults the most (Interaction Adult) and tend to be rated by the teacher as more dominant (TR Dominant). Factor five indicates that students who choose others as friends the most (Friendship Choices) report the least rule breaking in their environment (DOSE Rule Breaking). Paradoxically, they tend to be the target of the most aggression (Receiving Aggression). Factor six shows that students who describe themselves as rule following (Self-Description of Rule Following) report more punishment in their stories (DOSE Punishment) and comply to commands of other children (Compliance Peer) more. They also are less verbally abusive to their peers (Non-Physical Aggression).

IV. SUMMARY OF CHANGES FROM PRE TO POST IN EACH CLASS

In the fourth week of treatment the experimental class teacher punished her students more often and obtained more compliance than during baseline. However, there was more rule breaking and verbal aggression during this week than baseline. During baseline she

punished most the children who commanded their peers. In the fourth week of treatment she punished most the children who worked the least. She also interacted most with them. This suggests she was deficient in reinforcing the students who did more work. On the other hand she became more accurate in judging the dominance and popularity of her students, and she rated them as more popular, responsible, and independent. Even though there was more rule breaking and non-physical aggression during the fourth week of treatment than during baseline, the students described more authority and less rule breaking in their stories at this time.

In the comparison class only two changes occurred between pretest and post test. In the post test, the students commanded each other more often and they complied to orders from an adult more often. (The decline in peer interaction is not meaningful since week one of baseline is lower than week four of treatment.)

CHAPTER IV

DISCUSSION

I. EFFECT OF INTERVENTION ON CLASS

The effect of the intervention on the experimental class was equivocal. In addition the presence of the substitute teacher in the experimental class during the post test week and the proximity of the end of the school year complicate interpretation of the data. Nevertheless several conclusions are justified.

We attempted to improve the classroom management skills of the experimental class teacher and were partially successful. She punished more and obtained more compliance. Her monitoring skills improved as shown in her increased accuracy in judging the dominance and popularity of her students. And she became more positive toward her students as reflected in her ratings of them. This increased accuracy and positivity of the experimental class teacher's ratings are important effects of the intervention since various teacher judgments have as much potential impact on a student's future as the actual skills he learns.

However, we failed in two respects. Firstly, the teacher appeared to adopt a negative reinforcement strategy, obtaining compliance through the threat of punishment instead of positively reinforcing desirable behaviors. Thus in terms of the teacher's behavior, contrary to expectation, the experimental class became more rather than less coercive. Secondly, although we can speculate that the decreased rule

breaking and increased authority found in the children's post test DOSE stories reflect an improved classroom atmosphere, this is not substantiated in the observational data. In particular, observed Rule Breaking and Non-Physical Aggression increased. The increase in Non-Physical Aggression appears to be related to the presence of a substitute teacher, but regardless still reflects a failure of the classroom management system.

II. HYPOTHESES AND FINDINGS RELATING TO INDIVIDUAL VARIABLES

Sociometric Measures

Dominance and Chosen as Friend. Both of these measures were strongly associated with the teachers' ratings in three of the four first factors. This suggests that teachers' attitudes toward their students are significantly influenced by the students' dominance and popularity. Because of the practical implications, this relationship should be explored further.

In two of the data sets (post-experimental and post-comparison) Dominance and Chosen as Friend are associated with each other. In the other two data sets they are independent of each other. This suggests that dominance and popularity are sometimes associated in a social system while at other times they are independent of each other. Chosen as Friend is not closely related to Friendship Choices and Mutual Choices.

Friendship Choices. In both classes Friendship Choices reflect students' attitudes toward their peers. In the experimental class in

pretest low Friendship Choices is associated with high Self-Description of Rule Following and with DOSE Rule Breaking. Perhaps these students have adopted the teacher's rules and separate themselves from the behavior problems in the class. On the other hand, in the experimental class post test, the students who chose a high number of friends on the sociometric test tend to not comply with mands from their peers.

Contrastingly, in the comparison class, children who make more friendship choices appear to be more submissive; in the pretest period these children command their peers (Mand Peer) less often, and during post test they are the objects of aggression (Receiving Aggression).

Also notice the inverse relationship between Friendship Choices and DOSE Rule Breaking in both the pre-experimental and post-comparison data sets. Apparently those who choose a lot of peers as friends do not want to describe them as rule breaking.

In summary, Friendship Choices relates to the other variables in contrasting fashion across time and setting. Each factor it loaded on suggested how a segment of the children related to their peers. As a "behavioroid" measure, Friendship Choices lends relatively high face validity to factors which would otherwise be ambiguous. For instance, in factor four of the pre-experimental data set, Friendship Choices enables us to interpret that some of the students divorce themselves from the chaotic behavior of a problem class. Thus Friendship Choices is a promising measure in tapping aspects of classroom group process.

Mutual Choices. Except in the pre-experimental data set, Mutual Choices appears to be more closely related to number of choices than

to Chosen as Friend. The relation of Mutual Choices and Receiving Aggression in the post-comparison class and its lack of relation to the teachers' ratings suggest it is not as accurate a measure of popularity as Chosen as Friend.

Teachers' Ratings

The most striking aspect of the teachers' ratings is that the factor analyses indicate they are the most central variables. However, they are highly intercorrelated and such a cluster of high intercorrelations can skew the analysis. Therefore they cannot be regarded as being especially central to the data sets. Secondly, the high intercorrelation of the dimensions indicate that their discriminant validity is low. Responsible and Independent are highly intercorrelated as are Dominant and Aggressive while Popular does appear to be more discriminative. In the post-experimental measurement in particular the "halo effect" is very strong. The teacher rated each student similarly on all dimensions. Her ratings refer to Dominance and Popularity more than to any other of the measured dimensions.

However, factor one pre-experimental, factor two post-comparison, and factor four post-comparison indicate the teachers' ratings are responsive to the students' behavior. In the pre-experimental class the teachers' ratings of Independent and Responsible relate to the students' work and lack of verbal aggression (Non-Physical Aggression) and Rule Breaking. (Solitary Non-Task behavior will be discussed in the section on observational measures.) In the post-comparison class the teachers' ratings Aggressive and Dominant relate to the students' observed Rule Breaking and Physical Aggression. Also in the

post-comparison class the teachers' rating Dominant is associated with students' commanding other students (Mand Peer) and interacting with the adult (Interaction Adult). The experimental class teacher became more accurate in her rating of popularity between pre and post test.

Finally the teachers' ratings are congruent with the self-report measures in the three factors on which they both load. In the pre-experimental class, ostensibly a problem class, the children whom the teacher rated as Dominant, Popular, and Aggressive took the least responsibility for following classroom rules (Wanted to). In the post-experimental class those children whom the teacher rated highest on all dimensions included more Authority in their descriptions of the social environment. In the pre-comparison class the children whom the teacher rated as most Responsible did take responsibility for rule breaking on the self-report measure (My fault).

In conclusion, the relations of the teachers' ratings to the other variables suggest the complexity of rating behavior. They referred to the students' behavior, particularly Work and the oppositional behaviors (Rule Breaking, Non-Physical Aggression, and Physical Aggression) and to a lesser extent their attitudes. In addition they appeared to be strongly influenced by the students' sociometric status, both Dominance and popularity (Chosen as Friend). This was true of all dimensions since the dimensions were highly intercorrelated.

Self-Report Measures

The self-report measures are consistently congruent with the other methods. In addition higher scores on all the self-report

measures except DOSE Aggression appear to be associated with acceptance of the classroom authority structure.

My fault/Wanted to. My fault loads with Solitary Non-Task behavior in the post-experimental class. (As I will discuss later, Solitary Non-Task behavior appears to reflect compliance to the classroom authority structure.) It loads with the teacher's rating Responsible in the pre-comparison class. Thus in both settings "My fault" is associated with diverse methods measuring responsibility. In the pre-experimental class, "Wanted to" loads negatively on the second factor with Dominance and teacher ratings of Dominant, Popular, and Aggressive. That is, students who reported they "had to" instead of "wanted to" follow rules were more dominant in the pre-experimental class. It loads positively with DOSE Punishment and Authority in the pre-comparison class. That is, the students who described more punishment and authority attributed more responsibility for rule following to themselves.

Self-Description of Rule Following. Self-Description of Rule Following appears to reflect not only acceptance of classroom rules but also isolation or submissiveness with peers. In the pre-experimental class it is associated with fewer Friendship Choices, in the pre-comparison class with less interaction with peers (Interaction Peer), and in the post-comparison class with compliance to peers (Compliance Peer).

DOSE Aggression. In the post-experimental class students who included more aggression in their stories reported taking less

responsibility for following the rules (Wanted to).

DOSE Rule Breaking. In two of the three factors on which it loads describing more Rule Breaking in DOSE stories appears to reflect greater compliance with the classroom rules (fourth factor pre-experimental group, third factor post-comparison group). Also, in the fourth factor of the pre-experimental group and fifth factor of post-comparison group DOSE Rule Breaking is inversely related to Friendship Choices. Perhaps wanting to have friends in the group prevents a child from "finking" by telling stories with rule breaking in them to an adult.

DOSE Punishment. DOSE Punishment appears to be associated with compliance with class rules. Inclusion of punishment in the stories is associated with the self-report measure of wanting to follow rules (Wanted to) in the pre-comparison class. In the post-comparison class it is associated with Self-Description of Rule Following, Compliance Peer, and lack of Non-Physical Aggression. In the pre-experimental group it is associated with less interaction with adults (Interaction Adult).

DOSE Authority. Inclusion of Authority in the stories is similar to inclusion of punishment in the pre-experimental and pre-comparison groups. In addition in the post-experimental group those students who included Authority in their stories were rated higher on all dimensions by the teacher and more dominant (Dominance) and popular by their peers (Chosen as Friend).

Observational Measures

Solitary Non-Task. In order to interpret this measure it is necessary to understand its inverse relationship with Work. This inverse relationship occurs since both are mutually exclusive high rate behaviors. (Work accounts for 30-50% and Solitary Non-Task behaviors 20-40% of all observed intervals.) Furthermore, both relate to solitary behavior. Thus if a child works a lot it is very unlikely that he will engage in Solitary Non-Task behavior. With these considerations in mind, one can discount the negative relationship between Solitary Non-Task behavior and two of the factors labeled Responsible (factor one pre-experimental and factor three post-comparison). Indeed Solitary Non-Task behavior is associated with two self-report measures of responsibility and rule following (My fault in factor six post-experimental and Self-Description of Rule Following in factor four pre-comparison).

The association of Solitary Non-Task behavior with these measures of responsibility suggests that independent activity is related to compliance with authority. This supports the Berland, Coe, Resch, and Wahler (1976) finding that independent toy play, a component of Solitary Non-Task behavior in Wahler's category system, is inversely related to oppositional behavior. In addition the level of Solitary Non-Task behavior is clearly higher in the comparison than in the experimental class (see Figure 1, page 21). One can speculate that a higher level contributes to a smoothly functioning class.

Work. In the experimental class, Work appears to strongly influence the teacher's ratings during the pretest. During the post test week, it is again associated with the teacher's response since at this time it is inversely related to punishment. In the pre and post-comparison class, Work is inversely associated with observed oppositional and aggressive behaviors (Rule Breaking, Non-Physical Aggression, and Physical Aggression).

Mand Peer. This behavior appears to be related both to dominance and aggression. It was salient to the teacher in the pre-experimental class; at that time the teacher appeared to punish this behavior more than at any other. In the post-comparison class, it is related to the teacher's rating Dominant. In the post-experimental class, it is associated with verbal aggression (Non-Physical Aggression).

Compliance Adult. In the post-comparison class, Compliance Adult is closely related to the oppositional factor. Presumably this is because the teacher more frequently commands the children who are oppositional.

Interaction Adult. In both the post-experimental and post-comparison classes, the adult appears to interact most with the oppositional children. This suggests a deficiency in positive reinforcement of appropriate behavior.

Aggression Non-Physical. In both classes, Non-Physical Aggression is a component of oppositional behavior. It is negatively related to Work and in the post-comparison class to self-report measures of

responsibility (Self-Description of Rule Following, DOSE Punishment).

Physical Aggression. In the comparison class both pre and post, Physical Aggression is a component of oppositional behavior. In the post-comparison class it is associated with the teacher's rating Aggressive and Dominant.

Receiving Aggression. In the post-comparison group, Receiving Aggression combines with high Friendship Choices to form a "scapegoat" dimension.

Rule Breaking. Rule Breaking negatively correlates with factor one "Responsible" in the pre-experimental class, and it is a component of factor five "Oppositional" in the pre-comparison group. Thus it is a component of an oppositional behavior cluster.

III. CONVERGENT VALIDITY OF DEPENDENT VARIABLES

The multi-measurement system developed and employed in this study appears to possess a promisingly high degree of convergent validity as seen both among divergent methods in the same data set and across occasions and settings. The cumulative pattern of interrelationships of the self-report data with data from other methods is easily interpretable and congruent. For instance, the observed oppositional behaviors (Rule Breaking, Non-Physical Aggression, Physical Aggression) clustered together and were negatively related to Work. The teachers' ratings were congruent with both the observed behaviors and with the self-report data. It should be noted, however, that congruence was

seen only in the overall pattern. High positive correlations of different methods of measuring the same thing were in general not stable across occasions and settings.

Unexpectedly, all the teachers' ratings were strongly related to the sociometric measures Dominance and Chosen as Friend. This relationship was stable across occasions and settings (with the exceptions of Chosen as Friend in the pre-experimental group and Dominance in the pre-comparison group). Perhaps they are related because teachers' ratings of students and students' sociometric choices are both social choices and thus are similar methods. Regardless, the close relationship between them suggests they are responsive to similar behaviors in the person being rated or chosen.

An apparent exception to the high convergent validity of the measurement system is the decrease in DOSE Rule Breaking in the experimental class while observed Rule Breaking simultaneously increased. However, several considerations suggest that DOSE Rule Breaking reflects a veridical change in the class. First, observed compliance to an adult (Compliance Adult) and observed Punishment increased in the experimental class. Secondly, the intervention staff anecdotally reported that the teacher became more effective. Third, observed Non-Physical Aggression increased just during the week of the substitute, i.e., during the post test week. Thus it is possible that the teacher was much more in control of her class and the students responded to this in their DOSE stories. Viewed from this perspective, DOSE Rule Breaking can be seen as referring to different behaviors than observed Rule Breaking.

Instruments designed to measure intensity of rule breaking and observations in other settings in the school could help answer this question. For instance, the teacher could keep a log of serious rule violations. DOSE Rule Breaking might correlate with the number of serious rule violations but not with the amount of observed Rule Breaking. Also, observations in the playground or recess areas may indicate that most rule breaking and the most serious rule breaking occurs there. Again one can speculate that DOSE Rule Breaking refers more to occurrences during unstructured periods than during academic periods.

IV. OVER-JUSTIFICATION EFFECT

Although not related to observed behaviors, the association of "Wanted to" with DOSE Punishment and DOSE Authority on factor six of the pre-comparison group has implications for the over-justification effect controversy. On this factor the children who take responsibility for following rules include the most punishment and authority in their stories. This contradicts self-perception theory which would predict that a child takes responsibility when he feels that he, not the adult, is controlling his behavior.

This relationship occurs in the pre-comparison class where the teacher punishes congruently with her ratings of the children. Perhaps the adult's effective control of the class encourages the children to feel in control of their own behavior. Of course, that is no more than speculation, but it does suggest that the context in which monitoring or external rewards are used must be considered. The use of a

multi-measure assessment such as utilized in this study can delimit conditions in which various classroom management techniques are effective and those in which they have undesirable side effects.

REFERENCES

REFERENCES

Acker, L. E., Oliver, P. R., Carmichael, J. A., and Ozerkevich, M. J. Interpersonal attractiveness and peer interactions during behavioural treatment of the target child. Canadian Journal of Behavioural Science, 1975, 7, 262-273.

Alden, S. E., Pettigrew, L. E., and Skiba, E. A. The effect of individual-contingent group reinforcement on popularity. Child Development, 1970, 41, 1191-1196.

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), Advances in experimental social psychology. Vol. 6. New York: Academic Press, 1972, 52-68.

Berland, R. M., Coe, T. D., Resch, E. E., and Wahler, R. G. Independent Toy Play: A Keystone Behavior for Oppositional Children. Paper presented at the meeting of the Association for the Advancement of Behavior Therapy, New York, December, 1976.

Brown, J. H. The differential effects of monitoring on achievement behavior. Journal of Educational Research, 1975, 68, 318.

Campbell, D. T., and Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Christy, P. R. Does use of tangible rewards with children affect peer observers? Journal of Applied Behavior Analysis, 1975, 8, 187-196.

Csapo, M. Peer models reverse the "one bad apple spoils the barrel" theory. Teaching Exceptional Children, 1972, 5, 20-24.

Deutsch, M., and Gerard, H. G. A study of normative and informational social and influence upon individual judgment. Journal of Abnormal and Social Psychology, 1955, 51, 629-636.

Drabman, R. S., and Lahey, B. B. Feedback in classroom behavior modification: Effects on the target and her classmates. Journal of Applied Behavior Analysis, 1974, 7, 591-598.

Drabman, R. S., Spitalnik, R., and Spitalnik, K. Sociometric and disruptive behavior as a function of four types of token reinforcement programs. Journal of Applied Behavior Analysis, 1974, 7, 93-101.

Felixbrod, J. J., and O'Leary, K. D. Effects of reinforcement on children's academic behavior as a function of self-determined and externally imposed contingencies. Journal of Applied Behavior Analysis, 1973, 6, 241-250.

Festinger, L., and Carlsmith, J. Cognitive consequences of forced compliance. Journal of Abnormal and Social Psychology, 1959, 58, 203-210.

French, J.R.P., and Raven, B. The basis of social power. In D. Cartwright (Ed.), Studies in social power. Ann Arbor: University of Michigan Press, 1959, 150-167.

Greene, D., and Lepper, M. R. Effects of extrinsic rewards on children's subsequent intrinsic interest. Child Development, 1974, 45, 1141-1145.

Huck, S. W., Cormier, W. H., and Bounds, W. G. Reading statistics and research. New York: Harper and Row, 1974.

Johnson, S. M., and Bolstad, O. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Eds.), Behavior change: Methodology, concepts, and practice. Champaign, Ill.: Research Press, 1973.

Kazdin, A. E., Silverman, N. A., and Sittler, J. L. The use of prompts to enhance vicarious effects of nonverbal approval. Journal of Applied Behavior Analysis, 1975, 8, 279-286.

Kelley, H. H., and Thibaut, J. W. Group problem solving. In G. Lindzey and C. Aronson (Eds.), The handbook of social psychology Vol. 4 (2nd ed.). Reading, Mass.: Addison-Wesley, 1969.

Kreuger, D. E. Operant group therapy with delinquent boys using therapists' versus peers' reinforcement. Dissertation Abstracts International, 1971, 31 (11-B).

Lepper, M. R., and Greene, D. On understanding overjustification: A reply to Reiss and Sushinsky. Journal of Personality and Social Psychology, 1976, 33, 25-35.

Lepper, M. R., Greene, D., and Nisbett, R. E. Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. Journal of Personality and Social Psychology, 1973, 28, 129-137.

Lovitt, T. C., Lovitt, A. O., Eaton, M. D., and Kirkwood, M. The deceleration of inappropriate comments by a natural consequence. Journal of School Psychology, 1973, 11, 148-154.

Nisbett, R. E., and Valins, S. Perceiving the causes of one's own behavior. In E. E. Jones et al., Attribution: Perceiving the causes of behavior. Morristown, N. J.: General Learning Press, 1972.

Noonan, J. R., and Thiebeault, R. Primary prevention in Appalachian Kentucky: Peer reinforcement of classroom attendance. Journal of Community Psychology, 1974, 2, 260-264.

Patterson, G. R., and Reid, J. B. Reciprocity and coercion: Two facets of social symptoms. In C. Neuringer and J. Michael (Eds.), Behavior modification in clinical psychology. New York: Appleton-Century-Crofts, 1970, 133-177.

Reiss, S., and Sushinsky, L. W. The competing response hypothesis of decreased play effects: A reply to Lepper and Greene. Journal of Personality and Social Psychology, 1976, 33, 233-244.

Solomon, R. W., and Wahler, R. G. Peer reinforcement control of classroom problem behavior. Journal of Applied Behavior Analysis, 1973, 6, 49-56.

Wahler, R. G., House, A. E., and Stambaugh, E. E. Ecological assessment of child problem behavior. New York: Pergamon Press, 1975.

APPENDIX

APPENDIX

DEFINITIONS OF OBSERVATIONAL CATEGORIES

1. Solitary Non-Task. Any solitary behavior is recorded in this category with the exceptions of work, rule breaking, and being punished.

2. Work. This category is scored for on-task behaviors in school which are either required or strongly encouraged. Thus, this category requires a judgement based upon knowledge of the expectations for the subjects in any particular setting.

3. Mand Adult.
4. Mand Peer. This category is scored for any instance of command or instruction directed to an adult or a peer by the subject unless it is non-physical aggression or rule breaking.

5. Compliance Adult.
6. Compliance Peer. This category is scored for any instance of compliance during an interval with a mand given by an adult or a peer during an immediately preceding interval. Compliance with an on-going rule or expectation (e.g., starting on school or work) is not scored as compliance.

7. Interaction Adult.
8. Interaction Peer. This category is scored for any instance of the subject's approach to or interaction with an adult or a peer with the exceptions of mand or compliance behaviors, non-physical or physical aggression, receiving aggression, rule breaking and punishment, or work. This includes participation in games or other activities which have an element of cooperation or teamwork. If the subject initiates contact but his intent is not clarified within the five second interval, this is judged in context. For example, approaching the teacher with book in hand during a study period would most likely be scored work while approaching her on the playground would be scored interaction adult.

9. Aggression Non-Physical. This category is scored for any instance of the subject's intentionally acting in such a way as to be aversive to another child or adult excluding physically touching or harming that person. Accurate scoring must take into account situational norms. For instance, the observer must be familiar with which words in a setting are seen as friendly and which are seen as insults. The two aggression categories are scored even if the behavior also fits another category.

10. Aggression Physical. This category is scored for any instance of the subject's intentionally acting in such a way as to be aversive to another child or adult which includes physical contact.

11. Receiving Aggression. This category is scored for any instance of a child passively receiving either physical or non-physical aggression.

12. Rule Breaking. This category is scored for non-compliance with rules or instructions imposed by adults. Correct scoring requires knowledge of the teacher's rules which are in effect during the observation period.

13. Punishment. This category is scored for any instance of an aversive consequence imposed by a legitimate authority figure. This can be a task, restriction, physical punishment, etc.

VITA

Michael H. Schwartz was born in Chicago, Illinois on October 22, 1943. He attended elementary and high schools in Indianapolis, Indiana and was graduated from Broad Ripple High School in June 1961. He received a Bachelor of Arts degree in English from Goddard College, Plainfield, Vermont in June 1966.

He was a child care worker and supervisor of a residential center of the Judge Baker Guidance Center, Boston, Massachusetts from 1966 until 1972.

He entered the doctoral program in clinical psychology at the University of Tennessee, Knoxville in September 1972. He served his clinical internship at the Palo Alto Veterans Administration Hospital, Palo Alto, California from September 1975 until September 1976.

He received his Doctor of Philosophy degree in clinical psychology from the University of Tennessee, Knoxville in December 1978.

At that time he was employed as a staff psychologist at the Osawatomie State Hospital, Osawatomie, Kansas.

He is married to the former Susan Frances Meagher. They have three daughters.