DOCUMENT RESUME

ED 067 808 EC 050 099

| | |
|---|---|
| AUTHOR | Walker, Hill M.; And Others |
| TITLE | Generalization and Maintenance of Classroom Treatment Effects. |
| INSTITUTION | Oregon Univ., Eugene. Dept . of Psychology.; Oregon Univ., Eugene. Dept. of Special Education. |
| SPONS AGENCY | Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.; National Inst. of Mental Health (DHEW), Bethesda, Md. |
| PUB DATE | Aug 72 |
| NOTE | 44p. |
| EDRS PRICE | MF-$0.65 HC-$3.29 |
| DESCRIPTORS | *Behavior Change; *Behavior Problems; *Class Management; Emotionally Disturbed; *Exceptional Child Research; Family Problems; *Followup Studies; Operant Conditioning; Student Behavior |

ABSTRACT

Reported were two experiments which investigated, respectively, the maintenance of appropriate classroom behavior in children with behavior problems following treatment in an experimental classroom and cross situational consistency and generalization of treatment effects. In the first experiment followup performances of two groups of five subjects each were compared after treatment in a token economy. With one group additional procedures were implemented in the regular classroom to facilitate maintenance of their post treatment appropriate behavior. Treatment effects for subjects receiving treatment plus maintenance generalized to a much greater extent over the long term than did treatment effects for subjects who received only experimental treatment. In the second experiment five of the children were observed in family interactions in the home to determine whether the children, who were clearly deviant at school, were also behavior problems at home. Children and families examined after treatment demonstrated more child deviancy and parental negativeness than before treatment. It was thought that suppression of behaviors in the school setting may have caused an increase in the same behaviors in the home. (GW)

# GENERALIZATION AND MAINTENANCE OF CLASSROOM TREATMENT EFFECTS[1]

Hill M. Walker, Stephen M. Johnson,
and Hyman Hops

Department of Special Education
and
Department of Psychology
University of Oregon
Eugene, Oregon 97403

August, 1972

# GENERALIZATION AND MAINTENANCE OF CLASSROOM TREATMENT EFFECTS

Hill M. Walker, Stephen M. Johnson,
and Hyman Hops

The field of behavior modification has developed a powerful technology for changing human behavior (Ullmann & Krasner, 1965; Krasner & Ullmann, 1965; and Becker, 1971). Countless studies have demonstrated that behavior modification procedures can be applied successfully in such varied settings as clinics, mental hospitals, schools, homes, and residential facilities for delinquent or handicapped populations.

However, there has been an increasing recognition that such treatment effects tend to be specific to the setting(s) in which they are produced (O'Leary & Drabman, 1971; Skinrud, 1972). Researchers and change agents have experienced considerable difficulty in getting behavior changes to generalize to settings in which the intervention procedures have not been implemented. A number of studies have demonstrated that "unprogrammed" generalization of treatment effects to nontreatment settings is the exception rather than the rule (Wahler, 1969; Kuypers, Becker, & O'Leary, 1968; Meichenbaum, Bowers, & Ross, 1968; O'Leary, Becker, Evans, & Saudargas, 1969; Walker, Mattson, & Buckley, 1971).

Wahler (1969) argues that this lack of setting generality can be explained in terms of the environmental antecedents and consequences that operate in different settings. If there is a close "match" between the

setting events and contingencies in a child's home and school environments, for example, then his behavior in these two settings should show greater similarity. Conversely, if these setting events and contingencies are quite dissimilar, his behavior in the two settings might be expected to be less similar. Thus, if stimulus variables and/or reinforcement contingencies are altered in order to change behavior in one setting, one would not necessarily expect the changed behavior to generalize to other settings in which these variables have not been altered.

Wahler (1969) conducted a study in which he found this hypothesis to be confirmed. Two children were selected who exhibited deviant behavior in both the home and school setting. Intervention procedures were implemented in the home setting for both children. Results indicated that the behavior of both children improved in the home setting with no concomitant change in the appropriate behavior rates of either child in the school setting. Only when similar intervention procedures were later implemented for each child in his respective school environment did his behavior change in that setting. These data suggest the need for more intensive investigations of cross-situational behavioral consistency. The study also indicates a need for additional research on the generalization of treatment effects to settings other than those in which intervention procedures have been implemented.

A problem closely related to setting generality is the generalization of treatment effects over time, after formal treatment procedures have been withdrawn. The available data on this question indicate that treatment effects do not automatically maintain when treatment procedures are abruptly withdrawn (Walker, Mattson, & Buckley, 1971; Walker & Buckley,

1968; Birnbrauer, Wolf, Kidder, & Tague, 1965; Kuypers, Becker, & O'Leary, 1968; and Patterson, Shaw, & Ebner, 1969).

Unless gradual fading procedures or other scheduling techniques are employed in the treatment process, it seems unlikely that the level of behavior change achieved during treatment would be maintained following intervention. A fading procedure could act to change the environmental stimuli so that they approximate more closely those conditions that will prevail following the end of formal treatment. Thus, if treatment procedures are implemented in one setting, it seems unlikely that (a) the resulting treatment effects would necessarily generalize to other settings (during treatment) in which the environmental setting events and contingencies had not been similarly altered and (b) that behavior changes due to treatment would be maintained in other dissimilar settings after intervention had abruptly ceased.

The present study investigated four questions related to behavioral consistency across settings and to generalization and maintenance of treatment effects both across time and settings. These were:

1. To what extent does behavior change produced in a highly structured environment such as an experimental classroom, generalize and maintain in a regular classroom setting following treatment?
2. Will maintenance of classroom treatment effects be greater, over the long term, for a group of subjects who receive treatment in an experimental classroom plus follow-up maintenance in their regular classrooms after treatment than for an equivalent group of subjects who only receive treatment in the experimental classroom?

3. To what extent do deviant children exhibit disruptive behavior across settings? That is, to what extent does a child who is deviant in one setting, such as the school, exhibit deviant behavior in another setting, such as the home?
4. To what extent will intervention procedures implemented in one setting be reflected in changed behavior in a second setting where there has been no intervention?

The present study consisted of two experiments. Experiment I investigated questions 1 and 2 while experiment II investigated questions 3 and 4.

## Experiment I

### Maintenance of Appropriate Classroom Behavior Following Treatment in an Experimental Classroom

The purpose of experiment I was to compare the follow-up performance of two groups of subjects after treatment in a token economy. Both groups received approximately 4 months of intensive treatment in a special setting. Following treatment, all subjects were returned to the regular classrooms from which they were referred.

For group I subjects, additional procedures were implemented in the regular classroom to facilitate maintenance of their post-treatment appropriate behavior. No such procedures were implemented for group II subjects. Long term follow-up data were collected on both groups of subjects.

## Method

### Subjects

Two groups of five subjects were referred for treatment to an experimental classroom. The first group consisted of four boys and one girl; the second of five boys. The children ranged in age from 6 to 9 years and were enrolled in grades one, two, or three. One subject from each group moved away from the area before the present study was concluded. None of the eight remaining subjects came from the same elementary school.

Children were referred because of disruptive or deviant behavior occurring within the regular classroom setting. All subjects were screened using behavior checklist ratings, standardized individual intelligence tests (WISC: Stanford-Binet), achievement tests, standard auditory, visual, and general health tests and behavioral observations taken in the regular classroom.

The subjects selected met the following criteria: (1) high scores on the acting-out subscale of the Walker Problem Behavior Identification Checklist (Walker, 1970); (2) high rates on such observable behaviors as noisy, aggressive, movement around the room, inappropriate peer interaction, and nonattending; (3) average or above average scores on intelligence tests; (4) inadequate academic performance (educational deficits in the basic skills areas, for the two groups, ranged from 3 months to 1.5+ years); (5) no gross physical or sensory deficits; and (6) extremely low rates of appropriate behavior in the regular classroom setting relative to their peers.

Settings

The experimental classroom facilities were adjoining and affiliated with a public elementary school in the Eugene School District. There are 32 elementary schools in the district with an average teacher-child ratio of 1:24. The primary area for academic activities contained six double desks (approximately 20" x 45" work surface), the teacher's desk, shelves and tables for the display of high interest materials for science and art projects and a carpentry room with a variety of tools and wood. Adjacent rooms provided sink and table facilities and an observation area with a one-way mirror. Space was also available for individual testing, tutoring, and remedial instruction. A small isolation (time out) room containing a chair and desk adjoined the classroom.

Observation Recording System

A behavioral coding system was used to record the classroom behavior of the subjects and the social consequences provided by their teachers and/or peers. The system consists of 11 precisely defined subject behavior categories and 7 other categories which are potential consequences. The appropriate behavior categories were: (WK) work, (NO) group activity, (VO) vocalization, (PH) physical contact, and (MO) movement. Inappropriate behavior categories were: (WK) work, (NA) nonattending, (NY) noisy (nonverbal), (VO) vocalization, (PH) physical contact, and (MO) movement. The 7 consequence codes were as follows: (A) attention, (P) praise, (D) disapproval, (O) ignore, (C) compliance, (NC) noncompliance, and (PH) physical + or -. The subject codes were designated *a priori* as appropriate or inappropriate in a classroom environment. The percent appropriate

behavior was computed by dividing the total frequency of appropriate behavior by the total number of all behaviors observed.

The subjects' behaviors and consequences were recorded every 15 seconds for 6-minute observation periods. Observers were free to code as many subject behaviors and consequences as occurred during each interval. At the beginning of each 15-second interval, timers mounted in the observers' clipboards emitted an auditory "bleep" in an earphone worn by the observer and flashed a light at the top of the clipboard. This signalled the observer to record the behaviors on the appropriate line of the observation form until the next signal occurred.

Observations of the subjects were made in the regular classroom prior to enrollment in the experimental classroom, during treatment, and upon their return to the regular classroom. Observers were instructed to remain as inconspicuous as possible and not interact with experimental subjects in any way. Baseline data for each subject consisted of a minimum of 120 minutes of observation in the regular classroom over a 2-week period. Daily observations were recorded during the treatment phase and weekly observations during maintenance and follow-up.

Observer Training

Eleven graduate and undergraduate students in education and psychology, interested in working with handicapped children, served as observers during the treatment and maintenance phases of the study. Follow-up data was collected by a graduate research assistant. At the end of formal treatment in the experimental classroom, group I subjects continued to be observed for another 4 months in the regular classroom under maintenance

procedures. Both group I and II subjects were observed for approximately 4 months of follow-up during the next academic year.

At the start of training, each new observer was given a copy of an observation form and manual (Walker, 1971) to read and master. Once the observation instructions and code definitions were memorized to the satisfaction of the observer, he was brought into the observation facilities of the experimental classroom (which had a one-way mirror) to practice recording behavior. In addition, a videotape of a previous groups of subjects was used in the training process. A training observer worked with each new observer and took simultaneous recordings to check their reliability during the training sessions.

The training process required apprxoimately 1 week with 1-hour sessions per day. Generally, new observers spent 2 days practicing observations and 3 days checking reliabilities with the trainer. The follow-up observer was initially trained in the experimental classroom observation facilities. After achieving a specified criterion, the observer and trainer practiced recording behavior in the regular classroom setting and further checks on observer reliability taken there.

Observer Reliability

Observer reliability was calculated using the percent agreement method. Each 15-second interval was scored for the number of agreements and disagreements between pairs of observers. The total number of agreements was divided by the total number of behaviors recorded (agreements plus disagreements) to obtain the reliability coefficient. For an agreement to be scored in any interval, a rather stringent procedure was used; observers

were required to agree on the behavior being coded as well as the type of agent response that followed the behavior.

All observers were required to meet the criterion of .90 for five consecutive 6-minute observations before their data was acceptable. In addition, it was found that weekly reliability checks were required to maintain acceptable inter-observer agreement. The mean reliability for the maintenance period was .93 with individual reliabilities ranging from .90 to .97. For the follow-up observer, the mean reliability was .90 with a range of .70 to .98.

Procedures

Experimental treatment. Both groups of subjects received identical treatment within the experimental classroom. A complete description of the treatment program is contained in Walker, Fiegenbaum, and Hops (1971).

Briefly, treatment consisted of a token economy combined with an intensive, remedial instruction program in the basic skills areas of reading, mathematics, spelling and vocabulary. The three primary nonacademic treatment variables were token reinforcement, social reinforcement, and response cost (subtracting earned points).

Maintenance procedures. The post-treatment maintenance procedures for group I subjects consisted of a teacher training and feedback system designed to maintain both teacher and child behavior. Each child's regular teacher was trained in behavior modification techniques prior to the child's return to the classroom. The purpose of the training was to acquaint the teacher with principles of behavior modification so that she could knowingly reinforce and thereby maintain his appropriate classroom behavior. Special attention was also given to maintaining the teacher's behavior.

Each teacher was contacted approximately 1 month prior to the child's return from the experimental classroom to plan a smooth reintegration back into his regular classroom. Problems associated with maintaining treatment gains, achieved in the experimental classroom, were discussed as was the importance of the teacher's role in achieving behavior maintenance.

A contract was established between each teacher and the research project which specified roles each would play in programming behavior maintenance. The contract provided for: (1) training the teacher in behavior modification techniques, (2) weekly monitoring of her performance, and (3) reinforcement consequences contingent upon her performance.

Teacher training. The teacher agreed to read and master a semi-programmed text entitled, Modifying Classroom Behavior (Buckley and Walker, 1970). The text deals with basic principles of behavior modification and the application of these principles in the modification of classroom behavior. Each teacher agreed to take a review test over the text and achieve a passing score of 90 percent correct. If the teacher did not achieve this criterion on the first try, she reread the text and retook the test until the criterion was met.

Monitoring of teacher behavior. Each teacher met once a week with a project staff member who acted as a supervisor and monitored the teacher's performance. The supervisor, a resource teacher, provided the regular class teacher with backup support, consultation, feedback, and additional training and supervision in the application of specific behavior modification techniques.

The supervisor did not suggest specific techniques for achieving behavior maintenance. It was the teacher's responsibility to select the

procedures and techniques she planned to use. Once selected, however, the supervisor provided as much support and guidance as possible in their implementation.

An observer also met weekly with each teacher and provided a graphic record of the child's percent appropriate behavior for each observation session. These data indicated how well the child's behavior was maintaining and was an indirect measure of the teacher's performance. In addition, the supervisor monitored these data carefully and discussed them with the teacher during weekly meetings.

Maintenance of teacher behavior. The contract between the project and each teacher provided for reinforcement of the teacher's behavior contingent upon her performance. If the teacher fulfilled the provisions of the contract, the research project arranged for her to receive 6 hours of University credit under the course title Ed 505: Classroom Management Procedures, and paid her tuition.

The teacher's grade was dependent upon how well the child's behavior maintained during the follow-up period for the remainder of the school year (approximately 4 months). The figures for each child were discussed with their respective teachers at the time the contract was signed. If the child maintained 85 percent or better of the average amount of appropriate behavior he produced during treatment in the experimental classroom, the teacher received an A grade. If he maintained between 85 percent and 75 percent of this figure, the teacher earned a B grade. A C grade was earned if the child maintained at 74 percent or less. Teachers were able to use the data as criteria in evaluating their own performance during follow-up.

Prior research (Walker, Mattson, & Buckley, 1971; Walker & Buckley, in press) indicates that if no maintenance procedures are implemented following treatment in a token economy, appropriate behavior will show a considerable decline upon reintegration into the regular classroom. This may be due, in part, to the response cost involved in the extra effort required by the teacher to achieve maintenance. The authors attempted to construct ratios between behavior rates in treatment and maintenance that would be reasonable in the requirements they placed upon teachers. However, it was hoped that the ratio requirements, coupled with appropriate reinforcement consequences, would be instrumental in achieving adequate behavior maintenance.

Group I subjects were enrolled in the experimental classroom from October through January of the 1970-71 school year; group II subjects from February through June of the same school year. The maintenance procedures for group I were used during the remainder of the academic year. Both groups were followed up for the first 4 months of the 1971-72 school year. No maintenance procedures were implemented for either group during this 4-month period. Thus, it was possible to compare the follow-up performance of group I subjects who received treatment plus maintenance with group II subjects who only received treatment.

## Results

The mean percent appropriate behavior during baseline was 33.83 percent for group I subjects (range = 24.05%-39.15%) and 40 percent for group II subjects (range = 35.23%-45.86%). Following an arcsin transformation of

the proportion scores, a t test was computed for the difference between the means and was found to be not statistically significant ($t = 1.27$, $df = 6$, $p$ = n.s.).

While both groups made significant gains during treatment, there was no significant difference between their mean percent appropriate behavior during the last two weeks of experimental treatment; for group I, it was 94.11 percent and 97.66 percent for group II ($t = 1.50$, $df = 6$, $p$ = n.s.). Thus, groups I and II did not differ in their levels of appropriate behavior during baseline in the regular classroom or at the end of intervention in the experimental classroom.

---

Insert Table 1 About Here

---

Table 1 contains the means and standard deviations of the percent appropriate behavior for group I subjects during baseline in their regular classrooms, during the last two weeks of intervention in the experimental classroom, and upon their return to their respective regular classrooms under the maintenance procedures. The means during these periods were 33.83 percent, 94.11 percent, and 87.27 percent, respectively. A repeated measures ANOVA of the arcsin transformed scores was statistically significant ($F = 197.27$, $df = 2/6$, $p < .001$). A Tukey post-hoc test for the difference between the means showed the baseline level to be significantly lower than both the end of treatment ($p < .01$) and maintenance ($p < .01$) levels, but no significant difference between the means for the last two weeks of intervention and the maintenance period. These data show quite clearly that behavior maintenance was achieved for all four subjects.

---

Insert Table 2 About Here

---

The means and standard deviations of the percent appropriate behavior for group I and II subjects during the 4-month follow-up period in the subsequent academic year are presented in Table 2. Group I subjects averaged 80.25 percent, individuals ranging from 74 to 89 percent. The mean for group II subjects was 64.75 percent appropriate behavior during the same time period, with individuals ranging from 55 to 80 percent. A t-test of the arcsin transformed scores was statistically significant ($t = 1.97$, $df = 6$, $p < .05$). Thus, appropriate behavior was maintained much better for group I than group II subjects.

During the 4-month follow-up period, 12 minutes (two, 6-minute periods) of observation data were collected on each subject's regular classroom peers.

---

Insert Figure 1 About Here

---

Figure 1 contains the mean percent appropriate behavior for each experimental subject and his respective peer group. The peer groups for group I subjects averaged 79.75 percent appropriate behavior; for group II, the peer group mean was 78 percent. A t-test of the difference between the means of the transformed scores was not statistically significant ($t = 1.72$, $df = 6$, $p$ = n.s.). These data indicate that there were no differences in the general level of appropriate behavior in the classroom environments of the two experimental groups.

To further explore the relationship between the subjects and their peer groups, ratios were computed between each subject's mean percent appropriate behavior during follow-up and the mean of his peer group (see Table 3). A ratio of 1.00 indicates identical levels of appropriate behavior for the experimental subject and his peer group. A ratio of less

than 1.00 indicates the subject is performing above the mean of his peer group; a ratio greater than 1.00 indicates performance below the peer group mean.

---

Insert Table 3 About Here

---

The ratio data, presented in Table 3, show that the mean ratio for group I subjects was .99, indicating behavior very similar to that of the peer group. In contrast, the mean ratio for group II subjects (1.19) indicates that their percent appropriate behavior was somewhat below that of their peer group. However, no significant differences were found between each of the experimental group means and that of their respective peer groups.

Overall, group II subjects produced substantially less appropriate behavior in the follow-up than group I subjects. While there was some indication that group II subjects were considerably below the mean of their peer group, in contrast to group I, this difference was found not to be significant.

## Discussion

The results of experiment I suggest that a superior, long-term maintenance effect may be achieved if experimental treatment is accompanied by maintenance procedures implemented in the regular classroom setting following treatment. Treatment effects for group I subjects, who received treatment plus maintenance, generalized to a much greater extent over the long term than did treatment effects for group II subjects who only received experimental treatment.

The specific process accounting for this effect was not evident from data collected in experiment II. For example, group I subjects were exposed to some type of behavioral intervention, in either an experimental or regular classroom setting, for an entire academic year. Group II subjects were exposed to only 4 months of such treatment. Thus, the superior maintenance effect for group I subjects could have been due simply to an increased length of exposure to behavioral intervention procedures. Presumably, a longer treatment period would provide for the acquisition of social and academic skills that would be more resistant to extinction than those acquired within a brief treatment period.

It may be that length of treatment as a determinant of the long-term persistence of treatment effects operates independently of treatment setting. If so, the relationship should hold regardless of whether treatment is administered within an experimental or regular classroom setting. Consequently, the same long-term persistence of treatment effects might have been achieved if group I subjects had remained in the experimental classroom for the entire academic year. However, this is an empirical question that cannot be answered within the context of the present experiment.

An alternative explanation holds that processes unique to the maintenance procedures implemented for the group I subjects accounted for their superior behavior maintenance over the long term. For example, the maintenance procedures served to establish discriminative stimulus control (by the regular classroom teacher) over each subject's increased appropriate behavior. This control may have generalized to the group I subjects' regular classroom teachers during the next academic year thereby facilitating

maintenance of their increased appropriate behavior. Such an effect would not have obtained for group II subjects since no maintenance procedures in the regular classroom were implemented for them.

A third explanation argues that the superior maintenance effect for group I subjects was due neither to the length of treatment nor to the maintenance procedures, but rather to the classrooms in which the subjects were placed during the 1971-72 school year. That is, group I subjects were placed in classrooms in which higher levels of appropriate behavior were consistently produced whereas group II subjects were in classrooms in which lower average levels were consistently produced. However, inspection of Fig. 1 indicates that the peer data for the two groups do not support this hypothesis. The peer groups for group I subjects averaged 79.75 percent appropriate behavior while the peer groups for group II subjects averaged 78 percent appropriate behavior. The mean difference was not statistically significant. Thus, the results appear to be independent of variables unique to the respective classrooms in which the two groups of subjects were placed during the 1971-72 school year.

Regardless of the variable(s) accounting for the differential maintenance effect, it was encouraging to note that the behavior of subjects in both groups was maintaining well above pretreatment, baseline levels. This held true for all subjects in both groups. During long-term follow-up (one year), group I subjects averaged 80.25 percent appropriate behavior compared with 33.83 percent during baseline. Group II subjects averaged 64.75 percent appropriate behavior in long-term follow-up (8 months) compared with 40 percent during baseline.

It is interesting to note that $S_2$ in group II maintained at the mean of the group I subjects (80 percent) during follow-up even though no maintenance procedures were implemented for him. The other three subjects maintained well below the mean of the group I subjects. The authors have no information that would account for the much greater behavior maintenance of $S_2$. It may be that the academic skills of $S_2$ were such that his level of appropriate behavior was maintained by such natural reinforcers as task completion, academic success, teacher and peer praise of same, increased knowledge, etc. Again, it was not possible to confirm or deny this hypothesis in the present study.

This result points up a limitation of the present study. There was a small number of subjects in both groups; thus placing restrictions upon generalizations of the results beyond the experimental subjects. Replication of the results of this study using a larger number of subjects in both groups would be required to document the effect of either an increased treatment period or post-treatment maintenance procedures in accounting for superior, long-term maintenance of appropriate behavior.

Additional research is needed to evaluate the effectiveness of different treatments in producing behavior maintenance over the long term. Also, very little data is available on the question of how long a given treatment has to be in effect before long-term maintenance can be achieved. Intervention procedures have traditionally been evaluated in terms of the magnitude and efficiency of behavior changes attributable to them. However, this may be an inappropriate criterion. It could be argued that the most effective treatment is the one that produces the most durable maintenance of appropriate behavior across settings and situations, across behaviors, and across time.

On the other hand, evidence is accumulating that treatment effects tend to be specific to the settings in which they are produced (O'Leary & Drabman, 1971; Wahler, 1969; Skinrud, 1972; Meichenbaum, Bowers, & Ross, 1968). It may be considerably easier to achieve generalization of treatment effects over time within the same setting than it is to achieve generalization of treatment effects across different settings, e.g., home versus school. This may be due to stimulus control provided by the social agent who is instrumental in changing the child's behavior and who remains in that setting and continues to interact with him after treatment is terminated. For example, a therapist may work closely with a teacher, parents, or peers in implementing a program to change a child's behavior. Consequently, these social agents are paired with a powerful treatment and can acquire conditioned reinforcing properties as a result. In addition, they may continue to apply treatment techniques and principles after intensive, formal treatment procedures have been withdrawn and the therapist has discontinued his involvement in the case.

It is conceivable that generalization of treatment effects across settings may be a function in part of the amount of similarity that exists in the behavioral response classes elicited by stimuli operating in those settings. Thus, there may be a greater probability of generalization from one classroom to another than there is from classroom to home or from classroom to playground.

There are a large number of questions that remain to be investigated regarding generalization and maintenance processes. A number of studies are beginning to investigate whether generalization occurs during treatment and whether maintenance occurs after treatment. Fewer studies have investigated specific techniques for achieving maintenance.

## Experiment II

### Cross-Situational Consistency and Generalization of Treatment Effects

The purpose of experiment II was to examine the related questions of cross-situational consistency in behavior and generalization of behavioral treatment effects. More specifically, were children in experiment I, who were clearly deviant at school, also behavior problems at home? And, did the successful classroom treatment program have any systematic effect on the children's behavior in the home?

### Method

The subjects for this experiment were the five children who made up group II in experiment I. As indicated earlier, post-treatment data could be gathered on only four of these five subjects.

#### Settings

The children were observed in their homes under the same conditions employed in earlier normative research on child behavior and family interaction (e.g., see Johnson, Wahl, Martin, & Johansson, 1972). Each child and his family were observed for three consecutive days prior to and after experimental classroom treatment. The observation periods were 45 minutes each day and occurred during the hour prior to the family's regular dinner time when all family members were usually present. The families were required to comply with the following rules: (a) all family members were required to be present; (b) all family members were required to remain in a specified two-room area; (c) no interactions with the observer were permitted; (d) the television set was not allowed on; and (e) no visitors

were permitted and incoming telephone calls were to be quickly terminated. Parents were instructed to try to behave as they would if no observers were present and to present as representative a picture of the family as possible.

Observation System

A revision of the observational coding system developed by Patterson, Ray, Shaw, and Cobb (1969) was employed. The revised system utilizes 35 distinct behavior categories to record all the behaviors of the target child and family members who interact with him. The system is designed for rapid sequential recording of the child's behavior, the responses of family members, the child's ensuing response, etc. For purposes of determining observer agreement, all interactions were coded in the framework of 30-second intervals, and each observer was equipped with a 30-second stopwatch and a signaling apparatus. Each 30-second interval was broken down into interaction blocks in which were recorded the child's behaviors and family members' response(s) to those behaviors. Each block could contain one or two child behaviors, and one or two responses from each coded family agent. No more than two individual family agents could be coded as responding in each block. Provision was made, however, for the circumstance in which all family members present responded in the same manner.

Child behaviors which were continuous and without changes in family response were recorded every 10 seconds. Otherwise, behavior interchanges were coded as they occurred. In general, between three and four interaction blocks were recorded during each 30-second period. As in previous home observation studies, 15 of the 35 behaviors have been designated as "deviant" for young children; the sum of the rates of these behaviors

comprise the child's deviant behavior score. The 15 behaviors are those which a sample of 146 parents of young children have described most consistently as deviant. These behaviors also tend to receive relatively high proportions of negative consequences from the social environment (Wahl, Johnson, Johansson, & Martin, 1972). Adkins and Johnson (1972) have recently shown that there is a strong relationship between the average parent's rating of a behavior as deviant and the average family's tendency to respond to it in a negative manner.

The behavior codes used here have also been categorized on an a priori basis as serving either a positive, negative or neutral antecedent or consequent function. These categories reflect the investigators' assumptions about the intended functions of these behaviors under most circumstances. The face validity of this categorization is enhanced by evidence that behaviors which are deviant and/or negative produce a relatively high proportion of negative consequences (Johnson, et al., 1972). Furthermore, deviant behaviors in children are more often set off by negative antecedents than by positive antecedents (Wahl, et al., 1972).

The two dependent variables for experiment II were the proportion of deviant child behavior and the proportion of parental negative consequences. The deviant behavior codes were as follows:

Demand Attention
Violation of Standing Command
Destructiveness
High Rate
Humiliate
Noncompliance
Physical Negative
Smart Talk
Tease
Tantrum
Whine
Yell
Threatening Command
Ignore
Negativism

Negative parental consequences were:

| | |
|---|---|
| Threatening Command | Physical Negative |
| Command Negative (Terminating) | Smart Talk |
| Cry | Tease |
| Violation of Standing Command | Tantrum |
| Disapprove | Whine |
| Destructiveness | Yell |
| Humiliate | Demand Attention |
| Ignore | |
| Noncompliance | |
| Negativism | |

## Observers

Observations were made by a group of young female research assistants. Considerable effort was taken to keep the observers uninformed regarding all aspects of the experiment. They were uninformed regarding the status of the family (i.e., treated or normal control), the treatment stage, and the purpose and hypotheses of the present study. A different observer was always employed for pre and post assessments for any given child. As is often the case, however, some observers were informed by family members that they were in some form of treatment. In only 6 of the 30 observations, however, were observers informed of both the status of the child and the treatment stage. These 6 observation sessions involved the post assessment for two of the five children. Thus, only 22 percent of the observations could be considered to be threatened by observer bias, and even in these two cases, the observers were uninformed regarding the baseline level of deviant behavior in the target child.

## Observer Agreement

Seven of the 27 observation periods involved in this study were calibrated for observer agreement with all involved families being subject to

calibration checks at least once. As in experiment I, an overall percent agreement figure was computed in each case. To count as an agreement both observers had to agree on the same behavior for the same agent in the same interaction block. By this standard, the mean observer agreement was .69. As has been noted elsewhere (Johnson & Bolstad, 1972; Wahl, et al., 1972), this common method of estimating agreement is often overly conservative and not entirely appropriate for purposes of most research questions including the present ones. It is presented here to indicate that the reliability is comparable to that obtained in other research using this code and to justify generalization of other observer agreement data from larger available samples. In previous research where the average observer agreement figure was in the 65 to 75 percent range, it was found that the correlation between two observers of the overall deviant behavior proportion was .80 for one day of observation (Johnson & Bolstad, 1972). Since the statistics presented in the present study are for 3 days of observation, the Spearman-Brown correction for attenuation shows the expected observer agreement correlation for this extended period to be .95. The corrected observer agreement for the parental negativeness proportion was .97. Similarly, the corrected median observer agreement correlation for individual behaviors was .91 for 3 days of observation. Thus, while the average moment to moment agreement of two observers was only .69, the summary statistics (e.g., total number of deviant behavior) have been found to be highly consistent between two observers.

Reliability

It is also of interest to establish the reliability, as distinct from the observer agreement, on the summary statistics. Previous research

indicates that for 5 days of observation the uncorrected split-half reliability of the deviant behavior proportion was .63. The reliability of the parental negativeness score was .83. The median reliability of the rates of the individual behaviors was .53. Thus, estimates of these scores were reasonably consistent across observation periods as well as across observers in one period.

## Results

The availability of normative behavioral data on children of the same age in the classroom and home now makes it possible to determine whether children who have abnormal behavioral difficulties in the school also exhibit high levels of problematic behavior in the home. Of these five children who were clearly deviant in the school setting, all but one appear to be within normal limits in their rates of deviant behavior in the home. The average deviant behavior score for a sample of 40 same-aged normal children was 14.08 per day (s.d. = 13.60) comprising ~~3.90~~ 3.09 percent (s.d. = 2.69 percent) of the total observed behavior. The overall proportions of deviant behavior observed in the present sample, prior to treatment, are presented in the first column of Table 4. Only one child exceeded the normal mean (3.09 percent) in the deviant direction by more

---

Insert Table 4 About Here

---

than one standard deviation. Furthermore, although the number in the current sample is small, the normal and the present "school deviant" samples are statistically equivalent ($t = > 1$). Additionally, no noticeable or

systematic differences were apparent in the analysis of the 15 individual deviant behavior codes.

Similarly, no apparent or significant differences appeared in the overall parental summary scores. The pretreatment parental negativeness scores are presented in column 4 of Table 4. No family in the sample exceeded the same age normative average of 3.09 percent parental negative behavior (s.d. = 2.05%). Parents in this sample were no more or less positive or negative to their children than were parents in the normal sample. One child in the sample certainly did exhibit abnormally high rates of deviancy but his parents did not show particularly high rates of negativeness.

The level of child deviancy in family number five is one of the highest observed in over 130 observations of both normal and referred children and families. Contrary to expectations, however, this child was not the most deviant in school. In the group of five, he was only the fourth most disruptive in school. There was no significant relationship between the degree of deviancy displayed in school and that displayed at home (Spearman Rank Order correlation = .00). The meaning of this low level of relationship is affected, however, by the fact that there was very low variability in both measures, and the n is quite small.

The second question of interest concerns the behavior changes from pre to post-treatment observed in the home. The total deviant behavior percents for the second assessment are presented in column 2 of Table 4. In all four cases where data could be obtained, there was some increase in the proportion of deviant behavior on the second assessment. After arcsin transformations of these proportion scores, a t-test for paired

observations indicated that this trend was significant ($t = 3.04$, $df = 6$, $p < .05$). The same trend was observed for the proportion of parental negative responding in that all parents exhibited more negativeness on the second assessment ($t = 2.50$, $df = 6$, $p < .05$). The child deviancy proportions remained within "normal limits" as defined by the normative sample data as did all but one of the parental negativeness scores (i.e., family four). Thus, the scores of child deviancy and parental negativeness were consistently higher after treatment. Although the school treatment program produced dramatic changes in the classroom, there is certainly no evidence of positive generalization to behavior in the home. Rather, there is a trend which is obviously subject to several interpretations, in the opposite direction.

## Discussion

The first and most obvious implication of this research is that children who exhibit high rates of deviant behavior in school do not necessarily show similar difficulties at home. Neither do their parents deliver higher than average rates of negative consequences for their behavior. In other research (Lobitz & Johnson, 1972), it has been found that children who are referred for psychological treatment receive significantly more negative parental consequences than do nonreferred children. Parental negativeness was found to be a very reliable discriminator between referred and nonreferred children.

Robert Wahler (personal communication) has also reported similar findings from his research on cross-situational consistency. More specifically, Wahler observed five children referred for school problems in both

home and school using the same coding system in both settings. None of the five children who demonstrated behavior problems at school exhibited similar difficulties at home. Furthermore, the parents of these children did not report the occurrence of the same kinds of behavior at home that were seen as problematic in school. None of the children were seen by their parents as being particularly problematic in any way at home.

The present results together with those reported by Wahler would seem to call into question the not infrequent practice of referring parents for counseling because of their child's school behavior problems. To the extent that the present cases are representative, there would be only a 20 percent chance that a child who is deviant at school will also be observed to be deviant at home. Furthermore, based on the meager available data on generalization of treatment effects (e.g., O'Leary & Drabman, 1971; Wahler, 1969; Skinrud, 1972; and Meichenbaum, Bowers, & Ross, 1968), it would not appear likely that improved behavior in relation to the family would have any necessary impact on the child's behavior in the classroom. It is possible, of course, to involve parents as agents in a treatment program centered around the child's school behavior. This use of parental resources would not seem inappropriate since the direct target of such involvement is school behavior and such programs would presumably involve a good deal of parent-teacher contact. The central point is, of course, that behavior problems should be dealt with in the setting in which they occur. Parents who have little trouble with their child at home will understandably be less motivated to seek treatment for their school problem children and may quickly perceive the doubtful value of intervention limited to the parent-child relationship.

Over the past two years, the second author has been in charge of a behavior modification training program for parents of problem children. Over that period, a record of treatment completion has been compiled by referral source. This record indicates that referrals initiated by the school are the most likely of all sources examined to result in early termination. Not infrequently, this termination has been done by mutual agreement between the parents and therapists with a referral for school treatment either in the regular classroom or in a special classroom of the type described in experiment I. In almost every case, the need for treatment was first perceived by school personnel and, with only a few exceptions, behavior problems at home were minimal. In a few cases there was evidence for behavioral difficulties at home but, even then, parental motivation for counseling was often negligible. As might be expected, the best referral sources, in terms of treatment completion, were self-referral and referral by a pediatrician.

The fact that all four children and families examined after treatment demonstrated more child deviancy and parental negativeness is of considerable interest. This finding is open to several interpretations. The most obvious is that some kind of "behavioral contrast" effect was operative here. As others have noted in animal research, when certain forms of behavior are suppressed in one setting, they may tend to increase in another setting where similar controls are not operative (e.g., Freeman, 1971; Terrance, 1966). An example of this effect may also be found in a treatment study by Meichenbaum, Bowers, and Ross (1968). In this study, institutionalized behavior problem adolescent girls were initially reinforced for appropriate classroom behavior during the afternoon but not

during the morning. In a reanalysis of these data, Skinrud (1972) has pointed out that 9 out of the 10 subjects increased in their proportion of inappropriate behavior during the morning hours when no reinforcement procedures were in effect. A t-test for paired observations indicated that this change was significant ($p < .05$, two tailed).

As has been noted, the rates of attentive behavior which were achieved in the experimental classroom of the present study were extremely high and far above the normative levels found in most regular classrooms. Such high levels of attentive behavior are, of course, incompatible with the emission of any degree of deviant or "acting out" behavior in the classroom. It is, therefore, possible that the obvious reduction of deviant behavior in the experimental classroom and the consistent increase in deviant behavior at home are causally related. It is possible that behavioral contrast effects may be more pronounced as the level of behavioral control increases in the intervention setting. In the present study, the control achieved in the experimental classroom was clearly higher than that achieved for most children in any regular classroom.

The behavioral contrast interpretation should not, however, be accepted uncritically. The present investigation is merely descriptive in that no control group was employed. It is quite possible that the present results could have been obtained without the classroom intervention. It is conceivable, for instance, that families were less reactive to being observed on the second occasion and, as a result, more willing to present a more negative image of their child and family. Recent research by Johnson and Lobitz (1972) indicates that parents are capable of manipulating the deviancy level of their children in response to instructions. It is of

interest to note that parental negativeness scores increased significantly in the Johnson and Lobitz (1972) study as parents tried to make their child appear more deviant. It is also possible, of course, that the time of year at which these assessments were taken had something to do with the observed increase. The pretreatment assessments were done in the winter when children are more accustomed to being confined indoors while the post assessments were conducted in the spring. Thus, a clear interpretation of this interesting result must await the completion of research now under way in which control children and families are being observed at the same intervals used for the treated children.

In spite of the lack of a control group, these results clearly replicate the Wahler (1969) descriptive study and they are consistent with Skinrud's (1972) findings of no positive generalization of improved behavior at home to similar improvements at school. The writers are fairly confident in the interpretation that the present results at least do not provide any evidence for positive generalization from school to home. This interpretation is attenuated a bit, however, by the fact that all four children involved showed rather low rates of deviant behavior at home in the first place. It is possible that the results might differ for children who were clearly deviant in both settings. The little evidence that is available from Wahler (1969) would suggest, however, that this lack of generalization is a consistent finding across children of varying levels of initial deviance.

If the present finding suggesting the operation of a "behavioral contrast" effect is replicated with larger samples and appropriate controls, it would have profound implications for behavior modification treatment programs. At this point, it is sufficient to take note of the phenomenon

in this sample and await the results of further research which is now in progress.

## References

Adkins, D. A., & Johnson, S. M. An empirical approach to identifying deviant behavior in children. Paper presented at the Western Psychological Association Convention, Portland, Oregon, April, 1972.

Becker, W. C. An empirical basis for change in education. Chicago: Science Research Associates, 1971.

Birnbrauer, J. S., Wolf, M. M., Kidder, J., & Tague, C. E. Classroom behavior of retarded pupils with token reinforcement. Journal of Experimental Child Psychology, 1965, 2, 219-235.

Buckley, N. K., & Walker, H. M. Modifying classroom behavior: A manual of procedures for classroom teachers. Champaign, Illinois: Research Press, Inc., 1970.

Freeman, B. J. Behavioral contrast: Reinforcement frequency or response suppression? Psychological Bulletin, 1971, 75, 347-356.

Johnson, S. M., & Bolstad, O. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. Paper presented at the Fourth Banff International Conference on Behavior Modification, March, 1972.

Johnson, S. M., & Lobitz, G. Demand characteristics in naturalistic observation. Unpublished manuscript, University of Oregon, Eugene, Oregon, 1972.

Johnson, S. M., Wahl, G., Martin, S., & Johansson, S. How deviant is the normal child: A behavioral analysis of the preschool child and his family. Advances in Behavior Therapy, in press.

Krasner, L., & Ullmann, L. Research in behavior modification. New York: Holt, Rinehart, and Winston, 1965.

Kuypers, D. S., Becker, W. C., & O'Leary, K. D. How to make a token system fail. Exceptional Children, 1968, 35, 101-109.

Lobitz, G., & Johnson, S. M. Normal versus deviant--Fact or fantasy? Paper presented at the Western Psychological Association Convention, Portland, Oregon, April, 1972.

Meichenbaum, D. H., Bowers, K. S., & Ross, R. R. Modification of classroom behavior of institutionalized female adolescent offenders. Behavior Research and Therapy, 1968, 6, 343-353.

O'Leary, K. D., Becker, W. C., Evans, M., & 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., & Drabman, R. Token reinforcement programs in the classroom. Psychological Bulletin, 1971, 75, 379-398.

Patterson, G. R., Ray, R. S., Shaw, D. A., & Cobb, J. A. Manual for coding family interactions, sixth revision, 1969. Available from ASIS National Auxiliary Publications Service, in care of CCM Information Service, Inc., 090 Third Avenue, New York, N. Y. 10022. Document #01234.

Patterson, G. R., Shaw, D., & Ebner, M. Teachers, peers, and parents as agents of change in the classroom. In F. A. M. Benson (Ed.), Modifying deviant social behaviors in various classroom settings. Monograph #1, Department of Special Education, University of Oregon, Eugene, Oregon, 1969, 13-47.

Skinrud, K. Generalization of treatment effects from home to school settings. Unpublished manuscript, Oregon Research Institute, Eugene, Oregon, 1972.

Terrance, H. S. Stimulus control. In W. K. Honig (Ed.), Operant behavior: Areas of research and application. New York: Appleton-Century-Crofts, 1966, 299-302, 317, 323.

Ullmann, L., & Krasner, L. Case studies in behavior modification. New York: Holt, Rinehart, & Winston, 1965.

Wahl, G., Johnson, S. M., Martin, S., & Johansson, S. An operant analysis of child-family interaction. Behavior Therapy, in press.

Wahler, R. G. Setting generality: Some specific and general effects of child behavior therapy. Journal of Applied Behavior Analysis, 1969, 2, 239-246.

Walker, H. M. The Walker problem behavior identification checklist. Western Psychological Services, 1970. 12031 Wilshire Blvd., Los Angeles, California.

Walker, H. M. Early identification and assessment of behaviorally handicapped children in the primary grades. Report No. 2, Center at Oregon for Research in the Behavioral Education of the Handicapped, University of Oregon, Eugene, Oregon, 1971.

Walker, H. M., & Buckley, N. K. The use of positive reinforcement in conditioning attending behavior. Journal of Applied Behavior Analysis, 1968, 1(3), 245-250.

Walker, H. M., & Buckley, N. K. Programming generalization and maintenance of treatment effects across time and across settings. Journal of Applied Behavior Analysis, in press.

Walker, H. M., Fiegenbaum, W. E., & Hops, H. Components analysis and systematic replication of a treatment model for modifying deviant classroom behavior. Report No. 5, Center at Oregon for Research in the Behavioral Education of the Handicapped, University of Oregon, Eugene, Oregon, 1971.

Walker, H. M., Mattson, R. H., & Buckley, N. K. The functional analysis of behavior within an experimental classroom setting. In W. C. Becker (Ed.), An empirical basis for change in education. Chicago: Science Research Associates, 1971, 236-263.

## Footnotes

1 The preparation of this manuscript and the research reported herein was supported by grants from the Bureau of the Handicapped, U. S. Office of Education and from the National Institute of Mental Health.

Table 1

Means and Standard Deviations of Percent Appropriate Behavior for Group I Subjects During Baseline and Follow-up

| Subjects | Baseline | | Last Two Weeks of Intervention | | Maintenance | |
|---|---|---|---|---|---|---|
| | $\bar{x}$ | s.d. | $\bar{x}$ | s.d. | $\bar{x}$ | s.d. |
| 1 | 33.90 | 15.22 | 95.48 | 5.81 | 86.35 | 19.25 |
| 2 | 24.05 | 21.32 | 92.36 | 8.39 | 86.41 | 12.88 |
| 3 | 39.15 | 16.44 | 89.91 | 14.18 | 84.56 | 19.84 |
| 4 | 38.23 | 16.53 | 98.69 | 1.84 | 91.66 | 8.63 |
| $\bar{x}$ Total | 33.83 | 17.38 | 94.11 | 7.56 | 87.25 | 15.15 |

Table 2

Means and Standard Deviations of Percent Appropriate Behavior for Group I and Group II Subjects During Follow-up

| Subjects | Group I | | Group II | |
|---|---|---|---|---|
| | $\bar{x}$ | s.d. | $\bar{x}$ | s.d. |
| 1 | 74.00 | 15.05 | 57.00 | 15.05 |
| 2 | 89.00 | 11.50 | 80.00 | 10.36 |
| 3 | 75.00 | 11.62 | 67.00 | 15.91 |
| 4 | 83.00 | 15.61 | 55.00 | 20.29 |
| $\bar{x}$ Total | 80.25 | 13.45 | 64.75 | 15.40 |

Table 3

Ratios of Post-Treatment Appropriate Behavior for Experimental Subjects Relative to Their Peer Groups

| Subjects | Group I | Group II |
|---|---|---|
| 1 | 1.05 | .89 |
| 2 | 1.04 | .95 |
| 3 | 1.00 | 1.46 |
| 4 | .88 | 1.49 |
| $\bar{x}$ Total | .99 | 1.19 |

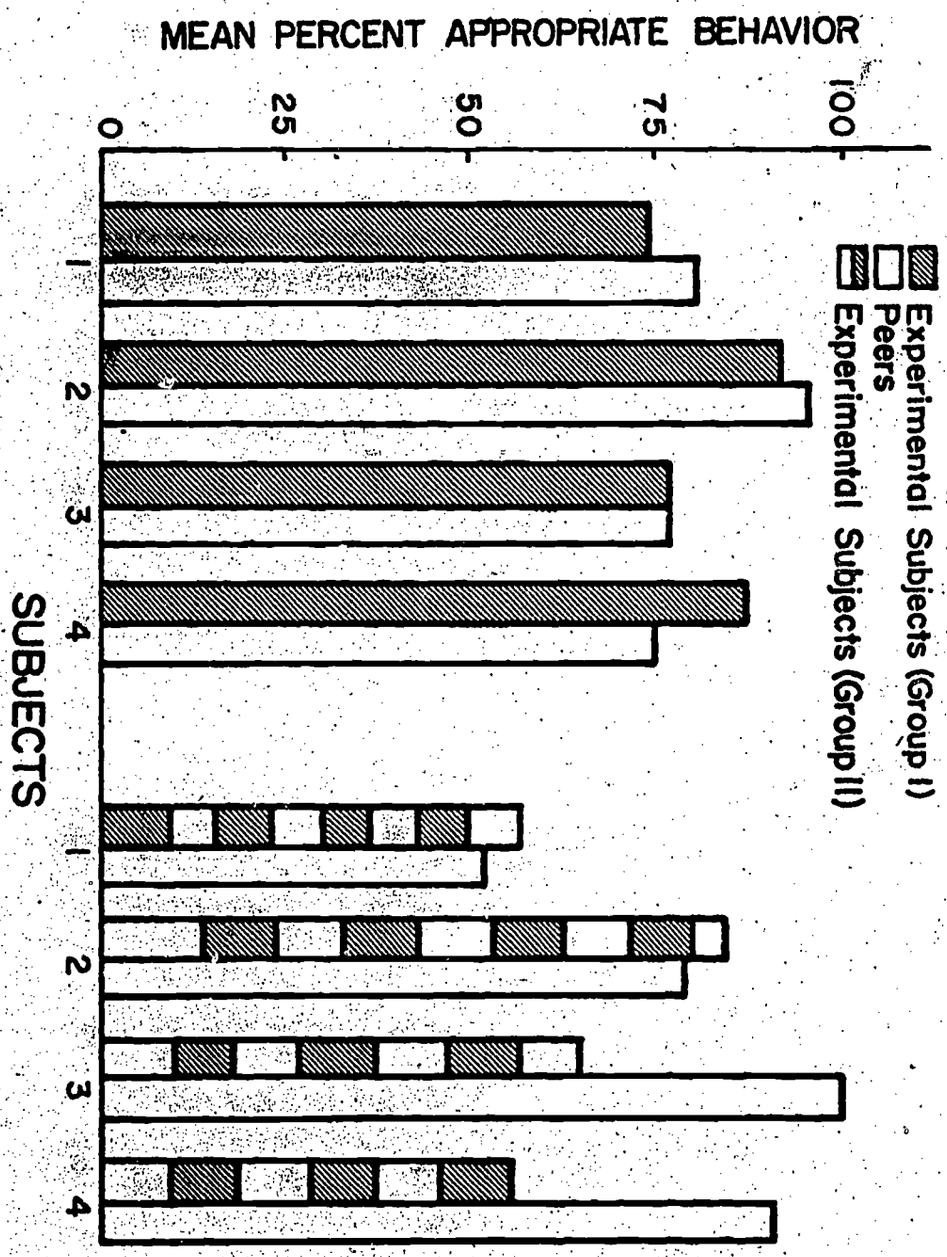
Table 4

Child Deviant Behavior and Parental Negativeness Scores Before and After School Treatment for Group II Subjects

| Family | Proportion: Child Deviant Behavior | | | Proportion: Parental Negativeness | | |
|---|---|---|---|---|---|---|
| | Pre | Post | Difference | Pre | Post | Difference |
| 1 | 1.5% | 4.1% | +2.6% | 0.4% | 1.9% | +1.5% |
| 2 | 2.0% | 4.5% | +2.5% | 2.1% | 3.3% | +0.8% |
| 3 | 0.7% | 1.1% | +0.4% | 1.3% | 1.8% | +0.5% |
| 4 | 1.2% | 1.7% | +0.5% | 2.0% | 7.6% | +5.6% |
| 5 | 14.1% | | | 2.5% | | |

## Figure Captions

Fig. 1 Mean Percent Appropriate Behavior for Experimental Subjects and Their Peers


MEAN PERCENT APPROPRIATE BEHAVIOR
0
25
50
75
100
SUBJECTS
1
2
3
4
1
2
3
4
Experimental Subjects (Group I)
Peers
Experimental Subjects (Group II)