


11-26-1975

# A Functional Analysis of the Effects of a Token Economy Program on Attending Behavior in Children and Subsequent Test Performance

Kay Delores Lillig Cotter  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)

 Part of the [Psychology Commons](#)

Let us know how access to this document benefits you.

---

## Recommended Citation

Lillig Cotter, Kay Delores, "A Functional Analysis of the Effects of a Token Economy Program on Attending Behavior in Children and Subsequent Test Performance" (1975). *Dissertations and Theses*. Paper 2206. <https://doi.org/10.15760/etd.2203>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

AN ABSTRACT OF THE THESIS OF Kay Delores Lillig Cotter for the Master of Science in Psychology presented November 26, 1975.

Title: A Functional Analysis of the Effects of a Token Economy Program on Attending Behavior in Children and Subsequent Test Performance.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

---

Gerald Guthrie, Chairman

---

James Paulson

---

Robert Jones

Thirty children, 4 years 8 months to 6 years 7 months, enrolled in a privately operated day school, were randomly assigned to either a treatment group or a control group over a one week period of baseline and a one week period of treatment. The experiment was designed to determine, if the implementation of a token economy program to decrease disruptive behavior in a classroom would result in improved academic performance. Both the children's and teacher's

behavior were assessed. The findings were nonsignificant for all of the dependent measures for children's behavior. The only significant effect for teacher's behavior was an increase in appropriate reinforcement behavior over time, for the control group. The study was prematurely terminated. The findings were discussed in terms of the practical and political implications of conducting research in a natural setting.

A FUNCTIONAL ANALYSIS OF THE EFFECTS OF A TOKEN ECONOMY  
PROGRAM ON ATTENDING BEHAVIOR IN CHILDREN  
AND SUBSEQUENT TEST PERFORMANCE

by

KAY DELORES LILLIG COTTER

A thesis submitted in partial fulfillment of the  
requirements for the degree of

MASTER OF SCIENCE  
in  
PSYCHOLOGY

Portland State University  
1975

TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the Committee approve the thesis of  
Kay Delores Lillig Cotter presented November 26, 1975.

Gerald Guthrie, Chairman

James Paulson

Robert Jones

APPROVED:

Robert Jones, Head, Department of Psychology

Richard Halley, Dean of Graduate Studies and Research

## ACKNOWLEDGMENTS

The writer wishes to express her sincere appreciation to her thesis advisor, Dr. Gerald Guthrie, for his assistance and support. She would also like to extend her sincere thanks to Dr. James Paulson and Dr. Robert Jones who served as members of her thesis committee. These three people were invaluable in helping her through some very trying times.

Thanks are also due to Ms. Diana Cady who served as the teacher, and Ms. Judith Sturmer and Ms. Dyan Harris who were observers. These people put in many hours of hard work to help the author in completion of her thesis.

The writer is deeply indebted to her husband, Dr. Patrick Cotter, for his help and support on the thesis. His comments and help on writing style were invaluable.

The writer wishes to extend her sincere gratitude to Ms. Candace Mueller for her help, friendship, and support during a difficult time. Ms. Sue Dixson also deserves thanks for her help and support.

Mr. Fred Sanchez deserves mention for his inconsiderateness and detrimental effect upon the author, the teachers and children of the day school where this thesis was conducted, and upon this thesis.

## TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS . . . . .	iii
LIST OF TABLES . . . . .	v
LIST OF FIGURES . . . . .	vi
INTRODUCTION . . . . .	1
METHOD . . . . .	7
Subjects . . . . .	7
Design . . . . .	7
Observation . . . . .	7
Procedure . . . . .	10
RESULTS . . . . .	15
Observer Reliability . . . . .	15
Observations of the Children . . . . .	15
Observations of the Teacher . . . . .	17
DISCUSSION . . . . .	22
REFERENCES . . . . .	33

LIST OF TABLES

TABLE	PAGE
I Means for the Children's Observations . . . . .	16
II Means for the Teacher's Observations. . . . .	19



LIST OF FIGURES

FIGURE	PAGE
1 Scoring Sheet for the Observations of the Children . . . . .	8
2 Scoring Sheet for the Observations of the Teacher. . . . .	9

## INTRODUCTION

The present experiment was designed to determine if the implementation of a token economy program to decrease disruptive behavior in a classroom would result in improved academic performance in children. The hypothesis proposed was that there is a relationship between quiet classroom behavior and academic performance. Both the children who were disruptive and those who attended to the disruptions had lost time during which they should have been attending to appropriate subject matter. If this inappropriate attending time were reduced, the children should have more time to attend to class work resulting in improved academic performance.

The token economy has been shown to be an effective controller of human behavior. It has been used with many types of target behaviors. It has been used to eliminate undesirable behaviors in psychotic patients (Atthowe & Krasner, 1968; Wincze, Leitenberg, & Agras, 1972). Token economies have been effectively used with retarded children, e.g. to increase school achievement (Dalton, Rubino, & Hislop, 1973) and with emotionally disturbed children to reduce latency of responding (Fjellstedt & Sulzer-Azaroff, 1973). Andrews and Ingham (1973) used a token economy in stuttering therapy. Increased housework behavior was

in nature. Fourteen third grade children, with a mean grade level of 2.58 on the arithmetic computation section of the Stanford Achievement Test, were administered a set of 100 arithmetic problems during a 20 minute drill, each day for 62 days. The problems were randomly selected from over 5000 arithmetic problems, testing arithmetic skills that had already been taught to the subjects.

During baseline the teacher circulated among the children to answer questions, but was told "not to attempt to accelerate attending behavior or suppress disruptive behavior." There were three treatment conditions. During the first treatment condition, the teacher administered tokens (plastic chips) for attending behavior, while ignoring non-attending behavior during testing.

Approximately once a week the tokens could be exchanged for candy, ice cream, toys, activities, or occasionally field trips. Treatment condition Two consisted of a token given for every seven problems worked correctly, plus bonus tokens for accuracy (one token for 21-30% correct, two tokens for 31-40% accuracy, up to eight tokens for 91-100%). The third treatment condition was a combination of Treatments One and Two, where both attending and correct work were reinforced. In order to correct for the increased number of tokens received in Treatment Three, the backup reinforcers were priced proportionately higher.

The children were observed on a 10 second basis. The

observer would watch one child for five seconds, then had five seconds to record his behavior. The observer scored each child in order of the class seating arrangement, and then started with the first child again.

The results were expressed in percentages, without the aid of sophisticated statistics. It was suggested from the results that

reinforcement contingencies for attending behavior alone were consistently effective in decreasing disruptions and increasing attending behavior but seemed to have little effect on correct work. When contingencies were placed solely on correct work, the accuracy increased but the number of problems correctly worked stayed constant, while attending behavior dropped and disruptions increased. Only when contingencies were placed simultaneously on attending behavior and on correct work did we find concurrent increases in attending behavior, number of problems worked correctly, and per cent of problems worked correctly (Ferritor, et. al., 1972).

This study included two experiments, with the second a follow up to the first. The second experiment was designed to correct for two methodological problems of the first experiment. The first problem involved the ordering of the treatment conditions. The authors hypothesized that perhaps the effects of Experiment One were due to an ordering effect. Therefore in Experiment Two, the order of the treatment conditions was changed to: Baseline, Treatment Two, Treatment One, Treatment Two, and Treatment Three. The second problem involved the increase in the number of tokens provided by the Treatment Three periods. The greater amount of tokens could have produced the effects found in the

Treatment Three periods. To control for the increase in number of tokens, the children received one token for every correct answer rather than one token for every seven correct answers in the Treatment Two conditions, plus the bonus points for accuracy given in Experiment One, consequently doubling the number of tokens received in the Treatment Two condition of Experiment Two. The other conditions remained the same, with Treatment Three being a combination of the Treatment One and new Treatment Two conditions. These changes did not effect the results. Experiment Two replicated the results of Experiment One.

Ferritor, et. al.'s (1972) experiment examined the effect of reinforcing attending behavior during testing on test performance. That is, the children were reinforced for attending behavior while they were taking the test. In addition, the children had already learned the arithmetic skills that were being tested. It is possible that the children had learned a finite amount of arithmetic and would continue to miss problems they did not know, regardless of whether they were attending during the testing situation. The data from the Ferritor, et. al. (1972) investigation would seem to support such a hypothesis. The results indicated that when the tokens were contingent on performance they did not increase the number of problems solved correctly, but the children stopped answering the problems they did not know, resulting in increased percent correct

scores. Consequently, they increased the number of tokens that they would obtain. During the final treatment period the mean number of correct responses did increase, but this could be due to some trial and error learning by the children, since they were informed of their errors and it was the last period providing several weeks for learning. More likely, it could be due to repetition of problems, since the problems were taken from a pool and then replaced in that pool to be used again. This would give the children a second chance on problems they had missed the first time.

In contrast to the Ferritor, et. al. (1972) study, the present investigation was designed to assess the effects of children attending to new information presented by the teacher and subsequently measuring the effects of that attending on test performance. That is, the child who has increased his attending to the teacher will have an increased probability of learning the new information being presented to him. The essential factor in this study, as in the Ferritor, et. al. (1972) study, was the class's test performance.

## METHOD

### Subjects

The subjects (Ss) were 30 children enrolled in the Towne Carousell Day School. Their ages ranged from 4 years 8 months to 6 years 7 months, with a mean age of 5 years 4 months. According to the school records, these children all fell within the normal range of intellect.

### Design

The experimental design consisted of an experimental group and a control group with pre- and post-measures for both groups. Both classrooms were observed four days per week for two weeks.

### Observation

Throughout the experiment, two observers (Os) were employed. One O recorded in both classrooms four days per week for the duration of the experiment. The second O recorded on randomly assigned days for a total of 4 days. For each observation checklist (children and teacher), the Os had 5 seconds to observe behavior and 10 seconds to record. The Os recorded the children's behavior for half of the total class time, while the remaining half was used to record teacher's behavior. The order of observation (teacher vs. children) was randomly determined on a pre-

arranged schedule. The children were assigned numbered seats. A random table of numbers was used to determine which child the Q started to observe first, each day. She then followed her observations of each child in order of the numbered seating arrangement. For the children's behavior, the scoring was done on a sheet containing three boxes for each observation. This layout was adapted from Ferritor, et.al., 1972, refer to Figure 1. The blank upper square was

	A		A
N	D	N	D
Ab	Ab		

Figure 1. Scoring sheet for the observations of the children.

for recording the seat number of the child observed. One of the three letters of the large upper square was "X"ed out. A represented attending behavior, N indicated nonattending behavior, and D represented disruptive behavior. The lower square was "X"ed, if the child was absent. The observation categories were as follows: (1) attending behaviors, which included a child looking at or writing on his paper, asking a question, looking toward the teacher, or passing out papers; (2) not attending, which included looking out the window, looking around the room, looking out the door, or sitting with closed eyes; (3) disrupting, which included verbal interruption of discussion, hitting another student,



making audible noises such as animal calls, moving the table around, drumming his pencil on the desk or wall, or throwing pencils, paper, or books. If disruptive behavior and attending or nonattending occurred during a five second interval, the O coded that time segment as a disruption.

The observation sheet for the teacher's observation session appears in Figure 2. The O "X"ed out one of the

R	A	R	A	
E	N	E	N	
P		P		

Figure 2. Scoring sheet for the observations of the teacher.

letters in the left column and one of the letters in the right of each square, after each five second observation. The O watched the teacher for five seconds, then had ten seconds to record the behavior, and continued in this manner for ten minutes. R indicated reinforcing behavior which included praise, talking, or yelling at a child in response to his behavior, smiling or jeering at a particular child, or describing behavior of a child or the class. E referred to extinction, which included the ignoring of the class, or not talking or cueing the class in any way. I stood for information. I included the teacher explaining the subject to the class, handing out materials, or other academically oriented activities where the teacher attended to the class.

P referred to time-out from positive reinforcement, e.g. placing a child against the wall or outside the room, or physical punishment of a child. A stood for appropriate behavior and N stood for inappropriate. R was defined as being appropriate for responding to the answer of a child, calling on a child with a raised hand, or praising the class. R was inappropriate for a verbal response to a child not in response to an answer or compliance to a command. E was defined as always appropriate. I was defined as appropriate as long as the discussion of one point did not exceed four minutes. P was defined as only appropriate for the use of time-out from positive reinforcement for a child's disruptive behavior and intercepting a child engaging in disruptive behavior, such as holding a child's arm to prevent him from hitting.

A second Q checked the reliability of the first at various times during the experiment. Both children and teacher observations were checked. The two Qs were trained before the experiment began to 79+% reliability, as determined by Scott's  $\kappa$  (Scott, 1955). Both Qs reviewed the observation procedures each week to insure the consistency of scoring.

### Procedure

Each class met 30 minutes each day. On Monday through Thursday, 20 minutes were spent on the teaching procedure. The teacher discussed one unit each week. Each day's lesson

consisted of five points related to the week's unit. An example of a week's unit is trees; the daily lesson, maple trees; the points 1) they mature fast, 2) they have broad leaves, 3) they shed their leaves every fall, 4) it's a soft wood, and 5) identification of the shape of their leaves. The teacher spent approximately four minutes on each point. The last five minutes of class time was spent administering a five point, yes-no, quiz with one question on each point of the lesson.

The teacher received standard instructions to follow through the entire experiment. They were reviewed by her on a weekly basis. The rules were as follows: 1) make all five points clear, by stating the point specifically and then discussing it; 2) give equal time to all points--approximately four minutes; 3) time-out from positive reinforcement is to be used for crying, yelling, physical aggression toward peers or the teacher, or destruction of property; 4) ignore all other disruptive behavior; 5) call on children who are attending, to ask or answer questions, and praise appropriate responses of those called on; and 6) ignore all other positive behavior.

Prior to baseline, the class was informed that they were going to receive tokens. That is, when they earned 288+ points (60% of the possible points) in four days, they would be able to vote on an activity which they could participate in on Friday, instead of their usual class. In

the event that they did not earn 288+ points, they would have class as usual. Activities included playing with balls outside, going jogging, going to watch the bowlers, playing with toys, games, or puzzles in class, playing in the recreation room, watching television, making popcorn, and any suggestion the children might have. The back-up reinforcers for the second week were a selection of toys, due to the fact that Friday was the Fourth of July.

At the end of each 60 second interval, an experimenter (E) located at the chalkboard administered tokens to the class by making slash marks on the board. During the first 20 minutes of class, the E administered either six, three, or no marks. The E recorded the number of tokens administered each time, on a scoring sheet. After the quiz was completed, the class was shown a graphic representation of the cumulative points they had acquired for the week. The graphic representation was a thermometer on 22" x 28" poster paper with an adjustable piece of red paper indicating points earned. The scale read from bottom to top with the minimum number of points needed to receive the back-up reinforcers at the top. During baseline the tokens were administered with six marks being administered 50%, three 20%, and none for 30% of the time. This insured that attending as well as nonattending and disruptive behaviors were reinforced. The class reached criterion with this schedule and received their back-up reinforcers on Friday.

The duration of the baseline condition was one week.

During treatment, the same token procedures were maintained, except that the number of tokens received was contingent on appropriate class behavior. Six points were given if three or less of the children were engaging in disruptive behavior, three points if four to six of the children were engaged in disruptive behavior, and no marks if more than six of the children were engaged in disruptive behavior. The percentage of tokens needed to reach criterion was increased by 5% the second week, based on the number of tokens that would have been earned during week one if tokens had been contingent. The increased percentage was used to effect a successive approximation toward attending behavior. The E reviewed the contingencies on a weekly basis. The treatment condition lasted one week.

The procedure for the control group was identical to the experimental group, except for the contingent token economy program. The experimental group preceded the control group each day. The control group received, noncontingently, the same pattern of token reinforcement that the treatment group received each day. That is, if the treatment group pattern of token reinforcement was 6, 3, 0, 3, 0, . . . on a particular day, the control group pattern of tokens would be 6, 3, 0, 3, 0, . . . regardless of the control group's behavior. This implied that possible teacher improvement would equally effect the treatment and

control groups.

## RESULTS

### Observer Reliability

Before the experiment began, the observer reliability was checked twice using Scott's  $\kappa$  (Scott, 1955). For the children's observations, the Scott's  $\kappa$ s were equal to 100% and 97%. The test for teacher's observations were equal to 95% and 79%. The first check was unduly high due to the observers exchanging comments during recording. Once the experiment began, two reliability checks were conducted each week for the two weeks of the experiment. The Scott's  $\kappa$ s for the children's observations were 66%, 90%, 78%, and 80% in order. For the teacher's observations, they were 88%, 68%, 82%, and 84% in chronological order.

### Observations of the Children

The dependent measures employed for the children's observations were the frequency of appropriate behavior, inappropriate behavior, disruptive behavior, absenteeism, and quiz scores. The experiment was ended prematurely. It was originally designed to last four weeks. Because it ran only two weeks, there was little hope of any significant results, so a scanning check was done to determine if  $t$  tests would be useful. The data analysis consisted of calculating the means for both treatment and control groups

for each of the dependent measures. The means were calculated by summing the total number of observations for a given measure over each week. These totals were then divided by the sum total of children present over the week, resulting in the number of observations recorded per child, for each week. Table 1 contains the means for all of the dependent measures.

TABLE I  
MEANS FOR THE CHILDREN'S OBSERVATIONS

Dependent Measures	Week 1		Week 2	
	Treatment	Control	Treatment	Control
Appropriate Behavior	1.82	1.67	1.49	1.56
Inappropriate Behavior	0.83	0.90	0.98	0.91
Disruptive Behavior	0.08	0.00	0.10	0.15
Absenteeism Behavior	1.25	0.67	1.37	0.85
Quiz Scores	2.98	3.39	2.80	2.48

The means for appropriate behavior were treatment group, week one 1.82; treatment group, week two 1.49; control group, week one 1.67; and control group, week two 1.56. It appeared that the groups were similar in respect to the amount of appropriate behavior. It also appeared that both groups remained constant over the weeks.



The means for inappropriate behavior were treatment group, week one 0.83; treatment group, week two 0.98; control group, week one 0.90; and control group, week two 0.91. It appeared that there were no significant differences between weeks or between groups for the children's inappropriate behavior.

For disruptive behavior, the means for the treatment group for weeks one and two were 0.08 and 0.10, respectively. For the control group, the means for weeks one and two were 0.00 and 0.15, respectively. Again, there appeared to be no significant differences between treatment and control groups, or for weeks of treatment.

The means for absenteeism for the treatment group were, week one 1.25 and week two 1.37. The means for the control group on this measure were, week one 0.67 and week two 0.85. The groups seemed similar in the amount of absenteeism and did not appear to vary over the weeks.

The mean scores of correct responses on the quizzes were, for the treatment group, week one 2.98 and week two 2.80; while for the control group, week one was 3.39 and week two was 2.48. Although there was more variance between means for the number of correct responses, there appeared to be no significant differences between treatment and control groups, nor for the weeks of treatment.

#### Observations of the Teacher

The dependent measures for the teacher's observations

were the frequency of teacher's appropriate reinforcement, teacher's inappropriate reinforcement, teacher's appropriate extinction, teacher's inappropriate extinction, teacher's appropriate information, teacher's inappropriate information, teacher's appropriate punishment, and teacher's inappropriate punishment. For each of the dependent measures, the mean number of observations was calculated by dividing the total number of observations for one week by the total number of days in that week. In those instances where visual inspection of the data suggested possible significant differences,  $t$  tests were employed to determine if they were significant at the 0.05 level of confidence.

For the teacher's appropriate reinforcement behavior, the means for the treatment group from week one to week two were 10.50 and 3.75 respectively. These were analyzed employing a two-tailed  $t$  test for correlated measures. The results showed that the teacher's appropriate reinforcement behavior did not significantly decrease from week one to week two ( $t=2.43$ ;  $df=3$ ;  $p>0.05$ ). The means for the control were 9.00 for week one and 4.25 for week two. A two-tailed coorelated  $t$  test was employed, indicating that the teacher significantly decreased her appropriate reinforcement behavior from week one to week two ( $t=3.45$ ;  $df=3$ ;  $p<0.05$ ). The differences between treatment and control groups, regardless of the week of treatment, did not appear to be significant. The teacher's observation data is represented

in Table 2.

TABLE II  
MEANS FOR THE TEACHER'S OBSERVATIONS

Dependent Measures	Week 1		Week 2	
	Treatment	Control	Treatment	Control
Appropriate Reinforcement Behavior	10.50	9.00	3.75	4.25
Inappropriate Reinforcement Behavior	3.75	5.50	5.50	5.25
Appropriate Extinction Behavior	1.00	0.75	0.50	0.25
Inappropriate Extinction Behavior	0.00	0.00	0.00	0.00
Appropriate Information Behavior	24.50	24.50	30.25	30.00
Inappropriate Information Behavior	0.00	0.00	0.00	0.00
Appropriate Punishment Behavior	0.00	0.25	0.25	0.00
Inappropriate Punishment Behavior	0.00	0.00	0.00	0.00

The means for the second dependent measure, teacher's inappropriate reinforcement behavior, were 3.75 for the treatment group, week one and 5.50 for the treatment group, week two, where as, the means for the control group were 5.50 and 5.25 for weeks one and two, respectively. These

findings suggested that the teacher's inappropriate reinforcement behavior did not differ for either treatment vs. control or for the weeks of treatment.

With regard to the teacher's appropriate extinction behavior, for the treatment group, the mean for week one was 1.00 and the mean for week two was 0.50. For the control group, the mean for week one was 0.75 and for week two the mean was 0.25. These results showed that the teacher's appropriate extinction behavior did not differ for either treatment vs. control or over time. For the teacher's inappropriate extinction behavior, all means were equal to zero, therefore there were no differences for treatment vs. control or over time.

Means for the teacher's appropriate information behavior were, for the treatment group, week one 24.50 and week two 30.25. The control group means were, for week one 24.50 and for week two 30.00. Upon visual inspection, a two-tailed  $t$  test was indicated. The results of the  $t$  test for the treatment group over weeks was not significant ( $t=1.79$ ;  $df=3$ ;  $p>0.05$ ). The results for the control group over weeks was, also, not significant ( $t=1.45$ ;  $df=3$ ;  $p>0.05$ ). Therefore, it can be stated that, for teacher's appropriate information behavior, there were no significant differences for either treatment vs. control groups or for week one vs. week two.

All of the means for teacher's inappropriate

information behavior were equal to zero. The teacher's appropriate punishment behavior means were 0.00 for the treatment group week one, 0.25 for the treatment group week two, 0.25 for the control group week one, and 0.00 for the control group week two. No significant differences were suggested upon inspection of this data, for either the treatment vs. control groups or over time. The final dependent measure was teacher's inappropriate punishment behavior, with means, again, all equal to zero.

## DISCUSSION

The original design of the experiment was to consist of one week of baseline and three weeks of treatment. After the completion of one week of baseline and one week of treatment, the experiment was prematurely terminated. The present discussion will focus on the events leading to this termination, the effect it had on the results, other interpretations of the results, and the problems of research in a natural setting.

One week prior to the initiation of the present investigation, the director was abruptly dismissed and immediately replaced by a new director. Subsequently, the new director made significant changes in the daily routine of the day school, resulting in both children and teacher dissatisfaction.

There is some anecdotal evidence of the children's dissatisfaction. The elementary school age children initiated a petition to be allowed to play in the "Big Room," the recreation room, after being told that the new director had discontinued use of this area. Several children expressed a desire to have the old director return. In addition, there was an increase in parents indicating that their children did not want to come to school, as well as, concern over their children's general unhappiness. Many of

the staff were approached by parents seeking different school placement for their children.

There was, also, anecdotal evidence of teacher dissatisfaction. Teachers were complaining about the changes being made and there was talk of resigning. Examples of changes that resulted in teacher dissatisfaction were: a) the elimination of use of the recreation room, b) forcing children to line up single file to change classes, c) the elementary school age children had to eat in total silence, while the other children napped, d) all children were required to be quieter than in the past, and e) one teacher had the same class of children all day, so there was no break for the teacher or the children. The new director, also, made promises he did not keep: a) reduction of class size, b) increased quality and quantity of food, c) getting needed supplies, e.g. crayons, paint, glue, and d) listening to teachers' suggestions. The above factors resulted in the teachers feeling a lot of stress and dissatisfaction.

After the director had been there three weeks and the experiment had ran two weeks, the director dismissed four of the staff. The staff dismissed included the experimenter, who was a teacher, and the experimental classroom teacher, resulting in the premature termination of the experiment. Looking at the results, there were several effects or lack of effects that could be attributed to the premature

termination of the study.

For the observations of the children, there were no significant results for any of the dependent measures. However, there was a trend away from the direction of hypothesized change for all measures. That is, the appropriate behavior and quiz scores decreased while the inappropriate, disruptive, and absenteeism behavior increased over time, regardless of group. This trend away from the hypothesized direction may be due to the children showing increased dissatisfaction with changes referred to earlier. The absence of significant results may be due to the short duration of the experiment. For one reason, the token economy might have been effective in increasing appropriate behavior if the study had run longer. The children may not have had time to acquire the secondary reinforcing effects of the tokens with only one week of treatment. Also, the new contingencies imposed upon the children and teacher by the director, resulting in the tensions and dissatisfaction discussed above, may have overridden any of the experimental contingencies, leaving only nonsignificant results.

The only significant result for the observations of the teacher was that the teacher's appropriate reinforcement behavior decreased over time for the control group. This finding may be due to an increase in teacher tension, making it less likely that she would give appropriate



reinforcement. The nonsignificant results for the observations of the teacher will be discussed later.

There are several findings which may be explained by variables other than mounting tension and the study's short duration. These results will now be examined. They include the reliability of test scores, the observations of the children, and the observations of the teacher.

The reliability test scores for the observers varied from 66% to 90%, while data was being collected. Although Scott's  $\kappa$  is a conservative test, the low scores suggest that there should have been more reliability between the observers in scoring. This could have been accomplished with more practice or perhaps better definitions.

As for the lack of significant results in the observations of the children, the token economy itself may have been ineffective, even if applied for a longer duration. Considering the young age of the subjects, individual tokens may have been more effective than group tokens.

The result that the teacher's appropriate reinforcement behavior significantly decreased over time for the control group may be due to the cumulative effects of the trends of the children's appropriate behavior decreasing and children's inappropriate behavior and disruptive behavior increasing. Therefore, the teacher had less opportunity to give appropriate reinforcement. There was a trend in the data indicating that the teacher increased the amount of

appropriate information given to the children. It may be that with increased information giving, there was less opportunity to use reinforcement for the children's appropriate behavior.

The remaining results for the observations of the teacher were all nonsignificant. These findings indicate that the teacher remained constant over time, following the rules given her at the onset of the experiment. It is important to note that, the lack of change in teacher's behavior over time was a desired effect, since any changes in children's behavior could then be attributed to the token economy system. Another possibility is that, if the children's behavior changed significantly in either direction, because of other variables, it may have resulted in changes or the lack of consistency in the teacher's behavior.

There are several alternative ways in which classroom investigations, using a token economy, may be conducted. Improvements may be made upon the system of data collection. While the system of data collection used in this study seemed adequate, the more measures that are taken the more data one has to work with. This becomes especially important when data collection is prematurely terminated. The extension of measurement could include obtaining duration measures for such behaviors as children's appropriate behavior, children's inappropriate behavior, and possibly children's disruptive behavior. Further improvement could

be made by obtaining more frequent measures of each child's behavior during an observational period. The duration of the data collection each day could, also, be extended. A follow up examination might be included to determine maintenance of improved classroom behavior and retention of information by the children. A final suggestion for further research would be a replication of the present study, incorporating the above mentioned improvements, since this study did not answer the questions it sought to, due to its early ending.

A commentary regarding the implementation of a study in a natural setting seems appropriate. There are many problems an experimenter may run into when trying to design and administer a research project in a real-life setting. Schwartz, Myers, and Astrachan (1973) addressed themselves to the problem of designing follow up studies on patients after treatment had already ended. Heathers (1974) discussed the problem of asking pertinent questions in educational research. It appears that, in education, little research has been done on the actual effectiveness of various teaching techniques. The author of this paper tried to ask a pertinent question in the present investigation, but ran into a different kind of problem. The problem can be labeled political. Reppucci and Saunders (1974) addressed themselves to this type of problem.

In a natural setting, it is often difficult if not

impossible to actually conduct a research project. Reppucci and Saunders tried to set up a behavior modification program in an institution. They discussed eight of the most common problems they encountered, i.e. 1) institutional constraints, 2) external pressure, 3) language, 4) two populations, 5) limited resources, 6) labeling, 7) perceived inflexibility, and 8) compromise.

They refer first to institutional constraints, i.e. bureaucracy or red tape. The present author found an example of this problem when trying to schedule her experiment. When the experiment was designed, classes of children rotated to different rooms and the teachers each had their own room. The new director did not like this schedule so he decided the classes would stay in their own room with the same teacher all day. This meant that the experimental teacher would not have two classes to use for the experiment. To resolve this problem, another teacher agreed to change rooms with the experimental teacher for an half hour each day.

There was also the problem of external pressure from outside the school. There were several state laws which impinged on the design of the study. One is that food could not be used as a reinforcer, although using snacks as reinforcers might have been more effective than tokens. Another law limited class size and therefore the number of children who could participate in the study.

Language was a problem that was faced by the experimenter when dealing with the teacher and observers. Guidelines were set up to define such words as reinforcement, punishment, and information. This was an essential part of the methodology of this study. It is an important consideration for anyone doing research with the aid of other personnel.

When research involves someone other than the experimenter administering contingencies, Reppucci and Saunders (1974) referred to this as a "two populations" problem. In a two populations problem, the experimenter modifies the staff who modify the subjects. In the present investigation, the author was forced to rely on the teacher to administer most of the contingencies in the classroom. In this type of situation, the experimenter must modify the the mediator, who modifies the subjects. The experimenter, in the present investigation, found the teacher readily modifiable, but this is not always the case.

Of course, there is always a problem of limited resources. In an institution, funding is usually limited and therapeutic research must be designed within these limitations. The problem faced by the present investigator involved her own pocket book. The supplies purchased included reams of paper, pencils, poster board, construction paper, and a stop watch. The investigator also had to locate a cassette player, a tape, and a second stop watch.

This list does not include getting permission to use various school resources.

The sixth problem was labeling. Activities often bear value-laden labels which limit the scope of the activities. Such labels include educational, recreational, nutritional, and therapeutic. In the day school, snacks were labeled nutritional and recess was labeled recreational. Because of their labels, neither one could be withheld or given contingently. Therefore, class time had to be set aside to provide for backup reinforcers, making activities on Friday contingent upon the points earned during the week.

Perceived inflexibility has also been stated as a problem. This problem involves the experimenter trying to maintain the basic integrity of his design, while remaining flexible. The author was approached by the director of the day school two weeks after the experiment started. She was asked if she could continue her experiment with a new teacher, since the experimental teacher had been fired. The author found it impossible to maintain the basic integrity of the research with a new teacher. There is a point where the flexibility required in a natural setting destroys the validity of the results.

The experimenter may find himself being modified by the experimental environment, in a real-life setting. This is labeled the problem of compromise. As Reppucci and Saunders explain it, "The behavior modifier is permitted

entry to the setting only so long as he embraces certain of the values of this core group of setting personnel." (1974). The present investigator objected to several of the changes made by the new director and was removed from the setting as a consequence.

The problems of perceived inflexibility and compromise were instrumental in causing the demise of the present investigation. These problems and the other six are apparently important considerations that a researcher must examine before and during the administration of an experiment. They seem to point out important factors to the success or failure of an experiment in a natural setting.

Here are some suggestions for protecting the experimenter and his study. It would seem wise to have a written contract with the management of an establishment to insure adequate control of the experimental design. Such a contract should include agreements of specific §s to be utilized; hours, days of the week, and duration of the experiment; and specific teacher, setting, and equipment to be used. Consideration should be given to the stability of management. One method of determining the stability of the staff and management would include interviewing the parties concerned. From the present investigator's experience, it is impossible to anticipate all of the variables that may impinge upon the scientific collection of data in a natural setting. However, the more information an experimenter has

about the social-political variables of the organization the more able he will be to control for these variables.



## REFERENCES

- Andrews, G. & Ingham, R. J. Stuttering: an evaluation of follow-up procedures for syllable-timed speech/token system therapy. Journal of Communication Disorders, 1972, 5, 307-319.
- Atthowe, J. M. & Krasner, L. Preliminary report on the application of contingent reinforcement procedures (token economy) on a "chronic" psychiatric ward. Journal of Abnormal Psychology, 1968, 73, 37-43.
- Birnbrauer, J. S., Wolf, M. M., Kidder, J. D., & Tague, C. E. Classroom behavior of retarded pupils with token reinforcement. Journal of Experimental Child Psychology, 1965, 2, 219-235.
- Chapman, C. & Risley, T. R. Anti-litter procedures in an urban high-density area. Journal of Applied Behavior Analysis, 1974, 2, 377-383.
- Dalton, A. J., Rubino, C. A., & Hislop, M. W. Some effects of token rewards on school achievement of children with Down's Syndrome. Journal of Applied Behavior Analysis, 1973, 6, 251-259.
- Drabman, R., Spitalnik, R., & Spitalnik, K. Sociometric and disruptive behavior as a function of four types of token reinforcement programs. Journal of Applied Behavior Analysis, 1974, 2, 93-101.
- Ferritor, D. E., Buckholdt, D., Hamblin, R. L., & Smith, L. The noneffects of contingent reinforcement for attending behavior on work accomplished. Journal of Applied Behavior Analysis, 1972, 5, 7-17.
- Fjellstedt, M. & Sulzer-Azaroff, B. Reducing the latency of a child's responding to instructions by means of a token system. Journal of Applied Behavior Analysis, 1973, 6, 125-130.
- Glynn, E. L. & Thomas, J. D. Effects of cueing on self-control of classroom behavior. Journal of Applied Behavior Analysis, 1974, 7, 299-306.

- Greenwood, C. R., Hops, H., Delquadri, J., & Guild, J. Group contingencies for group consequences in classroom management: a further analysis. Journal of Applied Behavior Analysis, 1974, 7, 413-425.
- Hall, R. V., Axelrod, S., Tyler, L., Grief, E., Jones, F. C., & Robertson, R. Modification of behavior problems in the home with a parent as observer and experimenter. Journal of Applied Behavior Analysis, 1972, 5, 53-64.
- Heathers, G. Planned educational change in search of a research tradition. Viewpoints, 1974, 50(3), 9-24.
- McKenzie, H. S., Clark, M., Wolf, M. M., Kothera, R., & Benson, C. Behavior modification of children with learning disabilities using grades as tokens and allowances-as back up reinforcers. Exceptional Children, 1968, 34, 745-752.
- McLaughlin, T. F. & Malaby, J. Intrinsic reinforcers in a classroom token economy. Journal of Applied Behavior Analysis, 1972, 5, 263-270. (a)
- McLaughlin, T. F. & Malaby, J. Reducing and measuring inappropriate verbalizations in a token classroom. Journal of Applied Behavior Analysis, 1972, 5, 329-333. (b)
- O'Leary, K. D. & Becker, W. C. Behavior modification of an adjustment class: a token reinforcement program. Exceptional Children, 1967, 33, 637-642.
- Reppucci, N. D. & Saunders, J. T. Social psychology of behavior modification problems of implementation in natural settings. American Psychologist, 1974, 29, 649-660.
- Schwartz, C. C., Myers, J. K., & Astrachan, B. M. The outcome in psychiatric evaluation research. Archives of General Psychiatry, 1973, 29, 98-102.
- Scott, W. A. Reliability of content analysis: the case of nominal scale coding. Public Opinion Quarterly, 1955, 19, 321-325.
- Sulzbacher, S. I. & Houser, J. E. A tactic to eliminate disruptive behaviors in the classroom: group contingent consequences. American Journal of Mental Deficiency, 1968, 73, 88-90.

Thompson, M., Brassell, W. R., Persons, S., Tucker, R., & Rollins, H. Contingency management in the schools: how often and how well does it work? American Educational Research Journal, 1974, 11, 19-28.

Wincze, J. P., Leitenberg, H., & Agras, W. S. The effects of token reinforcement and feedback on the dilusional verbal behavior of chronic paranoid schizophrenics. Journal of Applied Behavior Analysis, 1972, 5, 247-262.

Wolf, M. M., Giles, D. K., & Hall, R. V. Experiments with token reinforcement in a remedial classroom. Behavior Research and Therapy, 1968, 6, 51-64.