REWARD VERSUS COST TOKEN SYSTEMS: AN ANALYSIS OF THE EFFECTS ON STUDENTS AND TEACHER

BRIAN A. IWATA AND JON S. BAILEY[1]

FLORIDA STATE UNIVERSITY

The effects of reward and cost token procedures on the social and academic behavior of two groups of elementary special-education students were assessed using a reversal design. Behavioral observations of three target subjects in each group revealed that both procedures were about equally effective in reducing rule violations and off-task behavior. Records kept on the daily arithmetic performance of all subjects showed that output doubled in both groups during the token phases, although accuracy remained unchanged. When students were allowed to choose either contingency, no pattern of preference was established. Small differences were found in teacher behavior: the reward procedure led to an increase in approval comments but cost procedures produced no changes in teacher behavior.

In recent years, the token economy has emerged as a powerful tool in behavior management (Kazdin and Bootzin, 1972; O'Leary and Drabman, 1971; Paul, 1969). Although the range of subjects, settings, target behaviors, dispensing and recording procedures, and type of back-up reinforcers has varied widely, the procedures appear to produce reliable and replicable changes in behavior. The majority of these studies has emphasized what might be called a "reward" system of reinforcement, i.e., the subjects are rewarded with tokens when they have met some preset criterion behavior; non-target behavior receives no tokens.

Although several experimenters have reported the successful use of token systems based on response cost in reducing undesirable behavior (Broden, Hall, Dunlap, and Clark, 1970; Burchard, 1967; Phillips, 1968; Phillips, Phillips, Fixsen, and Wolf, 1971; Weisberg, Lieberman, and Winter, 1970; Winkler, 1970), others, viewing fines or cost as similar to punishment, have felt that the use of these techniques may lead to deleterious "side effects", such as increased aggression and escape behavior, which have typically been associated with punishment (Azrin and Holz, 1966).

Two direct comparisons between reward and cost procedures have yielded mixed results. McLaughlin and Malaby (1972) found that contingent point gain for quiet behavior produced lower rates of inappropriate verbalizations than did contingent point loss for inappropriate verbalizations on the part of fifth and sixth graders. Kaufman and O'Leary (1972), on the other hand, found no differences between reward and cost procedures in a psychiatric hospital school setting. Their results showed similar decreases in disruptive behavior and increases in academic performance for both the reward and the cost groups.

One purpose of the present study was to compare the effects of reward and cost token procedures on student social and academic behavior. A second purpose was to look at preference by the students for either a reward or cost token

system. During one phase of the study, students were given a choice as to which contingency (reward or cost) they preferred. Finally, records were kept of the teacher's verbal approval and disapproval in order to detect possible differential effects on teacher behavior that may result from the use of the two procedures.

METHOD

Subjects and Setting

Fifteen elementary school students in a special-education class (mean age 10 yr, mean IQ 70), who, according to teacher reports, exhibited a moderate to high degree of off-task and disruptive behavior, served. Based on prebaseline observations and teacher recommendations, the students were divided into two groups of seven (Group R-C) and eight (Group C-R) students respectively, of about equal arithmetic ability. Three students from each group were selected as targets for concentrated data collection on the basis of prebaseline observations because they exhibited the greatest amounts of off-task behavior. Experimental sessions were conducted each morning over a three-month period during a 40-min math class while students were engaged in individual seat work.

Observation

Daily observations were conducted during the first 30 min of the math period. Each of two observers was responsible for recording the behavior of three target students plus the teacher's verbal interaction with those three students. Observations were made on a continuous 10-sec interval basis, with the exception of the last 10-sec interval of each minute, which was reserved for making written comments.

The two categories of student behavior employed in the present study were:

1. *Off-task:* Visual nonattention to one's materials for more than 2 sec, unless the student was either talking to the teacher (with permission) or had his hand raised above his head.

2. *Rule violation:* the violation of one or more of the following teacher-prepared rules. (a) We remain seated during math class. (b) We raise our hands to get help from the teacher. (c) We do not talk or disturb our neighbor. (d) We may go to the bathroom when no one else is using it.

Since the categories were not mutually exclusive, it was possible for both categories to be scored during any 10-sec interval. However, only one instance of each category was recorded for a student during an interval.

The two categories of teacher behavior observed were:

1. *Teacher approval:* any verbal form of individual or group approval that was audible to the observer.

2. *Teacher disapproval:* any verbal form of individual or group disapproval that was audible to the observer.

Reliability

To assess the accuracy of the observers' records, frequent reliability checks were made by an independent observer. Reliabilities were calculated for each of the four separate categories of behavior by dividing the total number of agreements by the total number of agreements plus disagreements. An agreement was scored if both observers recorded the same behavior within the same 10-sec interval. A disagreement was scored if one observer recorded a behavior and the other did not. Instances in which both observers agreed that no behavior occurred were not counted as agreements when computing reliabilities. The inclusion of such instances would result in spuriously high reliability, especially with regard to low-frequency behaviors.

Academic Measures

Arithmetic materials consisted of the Singer Math Series (Random House), which is composed of several kits, each containing a number of problem cards, 20 problems per card. Daily records were kept of all 15 subjects' performance during math period. After each session, the students' papers were graded either by the teacher

or by an observer, and daily performance was recorded in terms of both the number of problems completed and the percentage correct.

Experimental Procedures

Prebaseline. Observations were begun during math period four weeks before formal data were collected. During this time, the teacher was not given any instruction regarding the enforcement of rules and was merely told to operate in her usual manner.

Baseline I. During this phase, an apparatus consisting of a cassette-taped series of short-duration auditory tone signals was introduced to familiarize the class with it. The apparatus divided the 40-min math class into 10 intervals ranging from 3 to 5 min in duration. One signal was delivered at the beginning of the first interval and at the end of each of the first nine intervals. The end of the tenth interval was marked by four signals in rapid succession.

On the first day of baseline, the teacher reviewed the class rules and handed out writing materials and arithmetic folders that contained the math card on which each child was currently working. The students were then instructed to begin work at the sound of the first signal and to stop when they heard four signals in rapid succession. At the end of the math period, students engaged in a 15-min mid-morning break, for snacks and juice. Children who could afford to do so paid a dime; those who were on a free breakfast and lunch program were given a free snack.

Token I. On the first day of the token program (Session 27), the teacher followed the usual procedures of handing out folders and reviewing the class rules. In addition, a cup was placed on the table in front of each student. Students in Group R-C (reward-cost) received empty cups; students in Group C-R (cost-reward) received cups containing 10 tokens (obtained from Peabody Language Development Kit). The teacher then explained to the children that she wanted to see how well they could obey the class rules. Students in Group R-C were told

that they could earn 10 tokens for following the rules during math class. Students in Group C-R were told they could keep the 10 tokens already given to them if they followed the rules, but that they would lose them for breaking rules. All students were told that they had to have at least six tokens in their cups by the end of math class in order to earn a snack. The teacher also informed the students about "surprise days", on which the three or four students who had earned the most tokens since the last "surprise day" would be eligible for a special bonus. Since students did not know when "surprise days" were to be held, it was important that they try to accumulate as many tokens as possible from day to day.

Throughout the math period, the signal apparatus served as a cueing device for the teacher. At the sound of each signal, the teacher placed one token in the cup of each student in Group R-C who had not violated any rules during the previous interval, and withdrew one token from the cup of each student in Group C-R who had violated one or more rules during the previous interval. The final occasion for delivering or withdrawing tokens was indicated by the four signal series at the end of the math class.

At the end of the math period, observers counted the number of tokens in each student's cup and the experimenter recorded the daily totals on the bulletin board chart. All tokens and materials were then collected and snacks were served to those students who met the criteria. Students who were not eligible to receive a snack were permitted to engage in free-time activities of a neutral type (*e.g.*, stacking blocks, playing with an abacus) in a different section of the classroom.

On the sixth day of the first token phase (Session 32), the criterion for a snack was raised from six to eight tokens and remained at this level for the duration of Token I. "Surprise days" were held following Sessions 32 and 37 during the Token I phase. On the first surprise day, three students had achieved perfect scores (60 tokens). It was therefore decided that on

all subsequent surprise days, students with perfect scores would be eligible for a reward. The surprises consisted of a large variety of inexpensive toys and candy, ranging in price from 29 to 63 cents.

Baseline II. On the morning of the thirty-eighth session, the teacher announced that the token program would be discontinued for an indefinite amount of time due to a temporary inability to furnish free snacks and prizes. Procedures during this phase were the same as those followed during Baseline I.

Token II. The token program was re-instituted at Session 45. The procedures were similar to those used in the Token I phase, with the following exception. The group that previously earned tokens (Group R-C) was now the cost group and Group C-R became the reward group. The criterion for a snack was raised from six to eight tokens on the third day of this phase (Session 47), and one surprise day was held following Session 49.

Choice. Beginning at Session 55 and continuing for two additional sessions, each subject was given a daily choice between the reward and cost contingencies. On the basis of the students' choices, new reward and cost groups were formed daily.

RESULTS

Reliability

Reliability data were obtained for 40 of the 54 experimental sessions, yielding an overall mean of 87% and the following means for separate categories: rule violations—88%; off-task—86%; teacher approval—85%; teacher disapproval—90%. In addition to the 40 regular checks, four additional observations were made by a trained observer participating in another project who had no knowledge of the present study. These checks produced an overall mean of 92%, and the following category means: rule violations—84%; off-task—87%; teacher approval—98%; teacher disapproval—100%. Agreements on nonoccurrence of the behaviors were not included in these calculations.

Social Behavior

The mean percentages of intervals in which off-task behavior and rule violations were exhibited by the target subjects are presented in Figure 1, along with the group means for each experimental condition. Individual graphs were quite consistent with the group data. During Baseline I, the means for off-task behavior for Groups R-C and C-R were 30% and 32%, respectively; the mean for rule violations was 9% for both groups. With the institution of Token I, rule violations dropped to below 1% for both groups. Off-task behavior also decreased for both groups, especially after the increase in criterion from six to eight tokens. When the token system (Baseline II was withdrawn), off-task behavior increased for both groups to levels comparable to those of the previous baseline; only small increases in rule violations were observed. During the Token II phase, when the contingencies for Groups R-C and C-R were reversed, rule violations again decreased to 1% or below for both groups, and off-task behavior dropped to below 5% for both groups after the increase in criterion from six to eight tokens. During the final condition (Choice), each student was allowed to choose either the reward or the cost contingency. Although the main purpose of this phase was to assess preference, rather than the effectiveness of the reward and cost procedures, rule violations and off-task behavior continued at a low level (2% or less) for both the reward and cost groups.

Data on the net number of tokens earned by all subjects revealed Group R-C averages of 92 and 83 during the Token I and Token II phases, respectively, and Group C-R averages of 97 and 88 tokens. It appeared that Group C-R averaged more tokens than did Group R-C, and that reward and cost procedures had little or no effect on token earnings.

Arithmetic Performance

The mean numbers of problems completed per session and the percentage of problems cor-

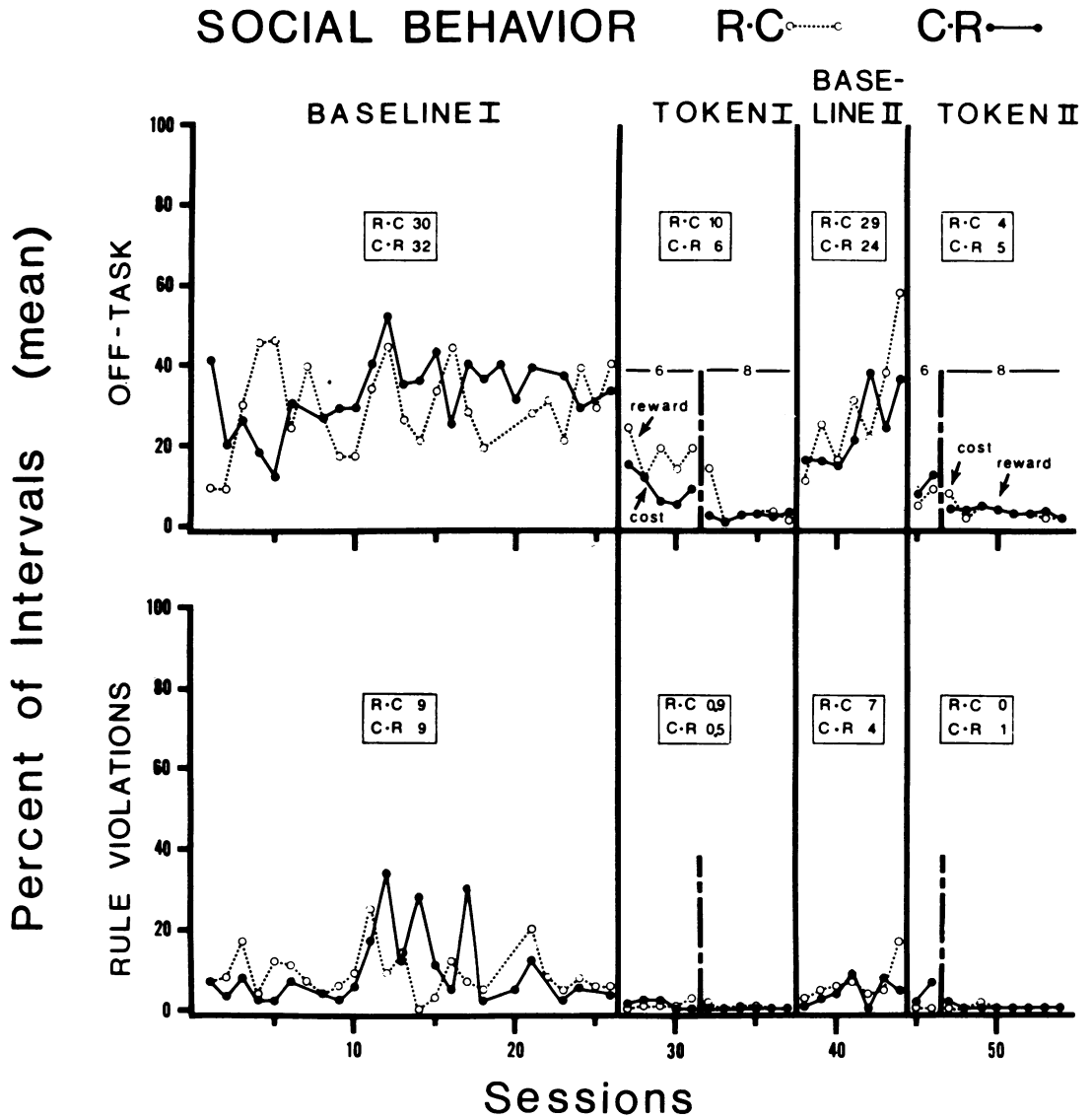## SOCIAL BEHAVIOR R·C o······c C·R •—•

Fig. 1. Mean per cent of intervals of off-task behavior and rule violations for Group R-C (Reward-Cost, N = 3) and Group C-R (Cost-Reward, N = 3). Group means for each condition are provided in boxes. The reward and cost groups during the token phases are indicated by arrows, and the change in criterion from six to eight tokens is indicated by a broken vertical line during Token I and Token II.

rect for all subjects in Groups R-C and C-R are presented in Figure 2. There was little difference between the groups during Baseline I in the number of problems completed. The means for Groups R-C and C-R during the first baseline were 19 and 21 problems, respectively. When Token I contingencies were put into effect, the output for Group C-R soon more than doubled, while the output for Group R-C increased only slightly at first. However, after the increase in token criterion, the performance of Group R-C rose to about the same level as Group C-R. When the contingencies were removed, arithmetic output decreased, and by the end of Baseline II, both groups had returned to Baseline I levels of performance. Output increased once again
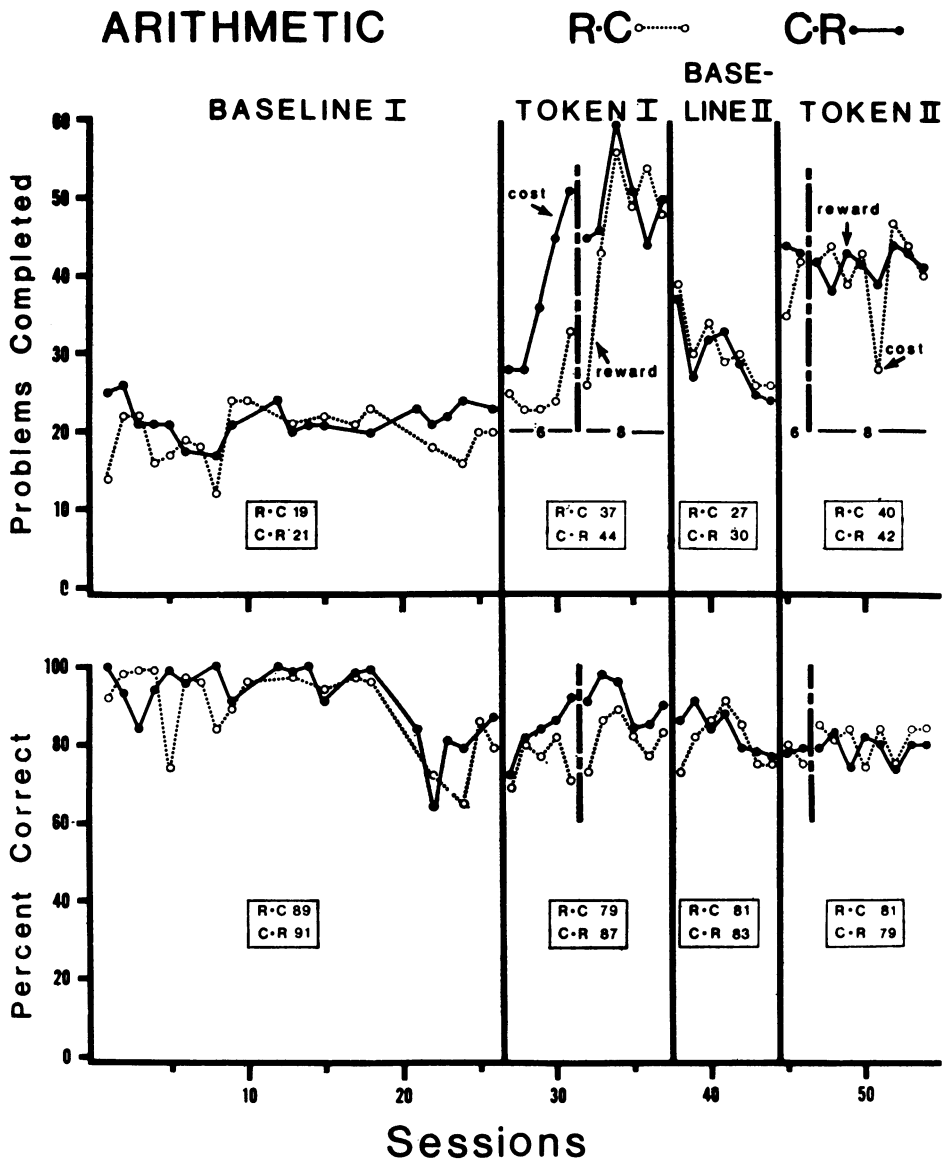
Fig. 2. Mean number of arithmetic problems completed and per cent correct for Group R-C (Reward-Cost, N = 7) and Group C-R (Cost-Reward, N = 8). Group means for each condition are provided in boxes. The reward and cost groups during the token phases are indicated by arrows, and the change in criterion from six to eight tokens is indicated by a broken vertical line during Token I and Token II. The nonrepetition and cheating control procedures were instituted in Session 20.

during the Token II phase and remained at a high level for both groups throughout the final Choice condition.

The data on the accuracy of problem solving revealed a drop for both groups, beginning with Session 20 during the Baseline I phase. Before this session, the teacher did not allow a student

to proceed to a new math card until he completed all problems on his present card correctly. Thus, some students were working on new cards, while others were repeating the same material. In addition, the observers began to notice an increase in cheating by the students (the answers to the problems could be found on the backs of the

math cards). Beginning with Session 20, students who completed a math card proceeded to a new card, regardless of their score. The teacher also clipped pieces of cardboard to the backs of the cards, so that the answers could not be seen.

As shown in Figure 2, reward and cost contingencies had little or no effect on the accuracy of arithmetic performance, which remained between about 80 and 90% for the entire study.

Teacher Behavior

The mean percentages of intervals during which the teacher delivered verbal approval and disapproval to the target subjects are presented in Figure 3. It can be seen that she delivered few comments of either a positive or negative type. Her Baseline I means for approvals and disapprovals were less than 2% of all observation intervals for both groups. Although she continued to maintain low levels of reinforcement throughout the study, there were small (less than 2%) but noticeable increases in approval for Group R-C during Token I (reward) and toward Group C-R during Token II (reward).

Student Choices

The final experimental condition was limited to three sessions due to the end of the school year. However, results of the 15 subjects' choices during this phase showed no consistent pattern
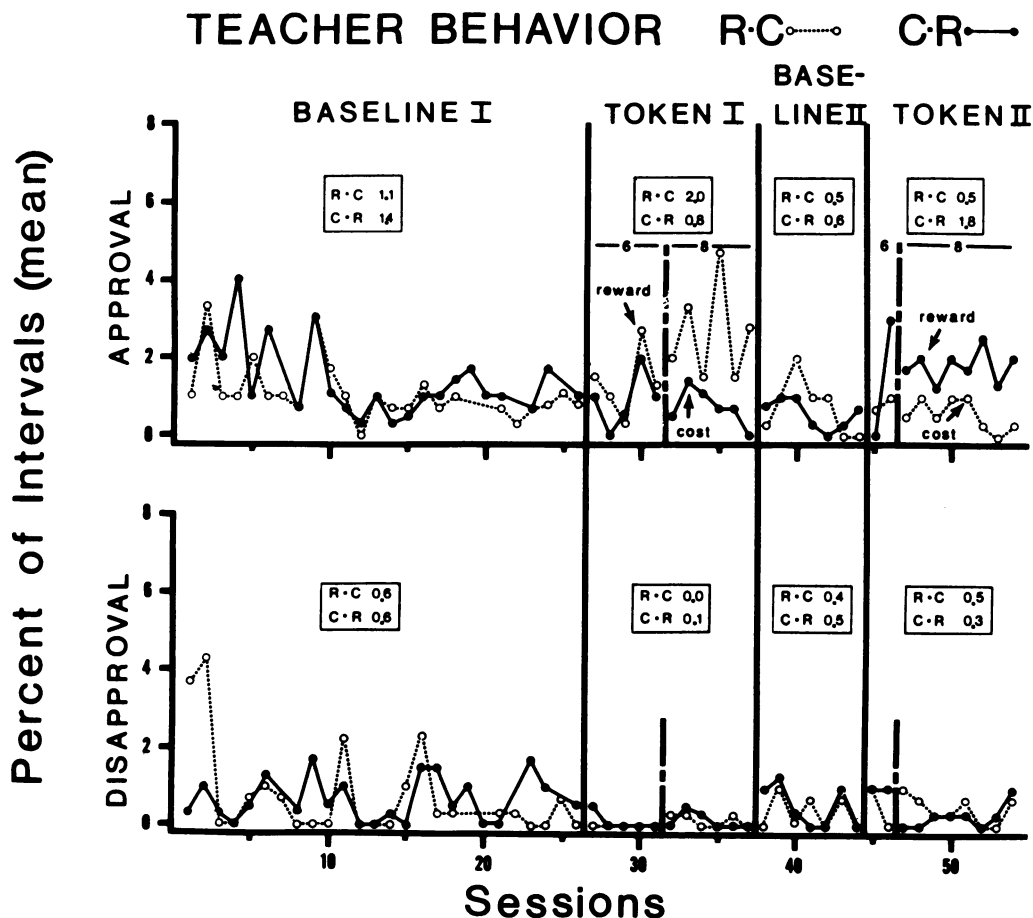

Fig. 3. Mean per cent of intervals of teacher approval and teacher disapproval for Group R-C (Reward-Cost, N = 3) and Group C-R (Cost-Reward, N = 3). Group means for each condition are provided in boxes. The reward and cost groups during the token phases are indicated by arrows, and the change in criteria from six to eight tokens is indicated by a broken vertical line during Token I and Token II.

of preference toward either reward or cost. Over the three-day period, four subjects consistently chose reward, five subjects consistently chose cost, and six subjects switched their choice at least once.

DISCUSSION

Present results indicated that both reward and response-cost token systems can be highly effective procedures in maintaining classroom social and academic behavior; in terms of student behavior, there appeared to be no differential effects resulting from the use of either contingency. Both procedures led to similar decreases in the percentage of classroom rule violations and off-task behavior of all students observed, even though the contingencies were not directly applied to the latter category.

A further indication of the success of the reward and response-cost procedures was the two-fold increase in academic output during the Token phases. Such a large change in behavior is perhaps also an indicator that academic output may be a more relevant measure of student behavior than the amount of time spent on-task, especially for students who behave appropriately most of the time. Three students who were not chosen for observation because of their good behavior, for example, went from Baseline I averages of 19, 22, and 32 problems completed per session to Token I averages of 64, 64, and 82 problems, respectively. The changes in arithmetic output are also representative of the beneficial "ripple effects" that may accrue from the application of contingency systems in general. Increases in the number of arithmetic problems completed were observed as a result of contingencies applied to classroom rule violations. Conversely, Ayllon and Roberts (1974) demonstrated that classroom disruptive behavior can be reduced by direct reinforcement of reading performance.

Perhaps the most meaningful measure of student behavior is actual academic achievement. Although the present procedures produced no change in arithmetic accuracy, the students were progressing at higher rates due to the fact that they were covering twice as much material in the same amount of time. A number of factors may have been responsible for the lack of gain in arithmetic accuracy. First, the present contingencies applied only to rule violations. Thus, students received no payoff for doing correct work. Second, when the teacher changed her former procedure of requiring students to answer all problems correctly before proceeding to a new card, the students may have taken this as a cue that it was more important to work rapidly than to work accurately. The institution of the Token I phase shortly after the teacher's change in procedures may have further led the students to believe that some type of contingency was being placed upon academic output. Finally, as the material became more difficult with each succeeding card, the students may not have had sufficient academic skills to progress satisfactorily without repeated instruction.

One of the major objections to the use of response cost has been that it may lead to detrimental side effects, such as increased aggression or behavior that enables one to avoid or escape the cost condition. The present results failed to support either of these contentions. Since aggressive behavior on the part of students was included in the rule violations category, increased aggression during the cost conditions would have appeared as an increase in rule violations for Group C-R during Token I, and an increase for Group R-C during Token II. Such increases did not occur. In fact, during both token phases, the cost groups engaged in fewer rule violations than the reward groups (See Figure 1). Absences from class also failed to establish cost as an aversive event. The combined absences for both groups under reward conditions averaged 0.48 per session, and absences under cost conditions were only slightly higher at 0.57 per session. Finally, data obtained from the Choice condition indicated that subjects found the cost contingency to be at least as "desirable" as the reward contingency, if not more so. Although more subjects

chose reward on the first day, cost was chosen more frequently on both of the following days.

In light of the conflicting results regarding the negative side effects of response cost, several factors or procedural differences may be associated with the presence or absence of these side effects. Kaufman and O'Leary (1972) mentioned a few, namely: the latency between the occurrence of the inappropriate behavior and the time at which the fine is levied, the value of the back-up reinforcers, the reinforcing value of the behavior that is being shaped, and the subjects' dislike toward the reinforcing agent. A further difference may be whether or not the tokens that are subtracted were initially given freely or earned contingently. The token dispensing procedures employed in both the present study and the Kaufman and O'Leary (1972) experiment were noncontingent, *i.e.*, students were given "free" tokens that were removed contingent on the occurrence of inappropriate behavior. On the other hand, Boren and Colman (1970) and McLaughlin and Malaby (1972), both of whom reported negative results in conjunction with response cost, employed a system in which the tokens removed had previously been earned contingent on the occurrence of appropriate behavior. Thus, a more punishing or aversive situation may be created when earned reinforcers are removed than when noncontingent reinforcers are removed.

A final point that deserves consideration regarding the relative merits of reward and cost is the effect that these procedures may have on the reinforcing agent. In the present study, the teacher delivered higher percentages of verbal approval toward the reward groups during the Token I and Token II phases. This effect was also replicated during the Choice condition. Thus, it appears that reward procedures may lead to slightly greater amounts of teacher approval than do cost procedures. There appeared to be no consistent indication that reward led to fewer disapprovals than response cost, or that response cost led to an increase in disapprovals over baseline rates. Although these results

should be considered highly tentative, due to both the teacher's low baseline of reinforcement and the exceedingly small changes in her behavior, it is conceivable that a teacher's behavior may be greatly affected as a result of implementing reward or cost procedures in the classroom. In turn, large changes in teacher behavior, especially in a negative direction, may have adverse effects on student behavior, the net effect being either that the teacher's verbal behavior undermines the positive effect of the tokens, or that the teacher must continuously rely on the token system for control in the absence of a system of social reinforcers toward which she can fade.

REFERENCES

Ayllon, T. and Roberts, M. D. Eliminating discipline problems by strengthening academic performance. *Journal of Applied Behavior Analysis,* 1974, **7,** 71-76.

Azrin, N. H. and Holz, W. C. Punishment. In W. K. Honig (Ed.), *Operant behavior: areas of research and application.* New York: Appleton-Century-Crofts, 1966. Pp. 380-447.

Boren, J. J. and Colman, A. D. Some experiments on reinforcement principles within a psychiatric ward for delinquent soldiers. *Journal of Applied Behavior Analysis,* 1970, **3,** 223-233.

Broden, M., Hall, R. V., Dunlap, A., and Clark, R. Effects of teacher attention and a token reinforcement system in a junior high school special education class. *Exceptional Children,* 1970, **36,** 341-349.

Burchard, J. D. Systematic socialization: a programmed environment for the habilitation of antisocial retardates. *Psychological Record,* 1967, **17,** 461-476.

Kaufman, K. F. and O'Leary, K. D. Reward, cost, and self-evaluation procedures for disruptive adolescents in a psychiatric hospital school. *Journal of Applied Behavior Analysis,* 1972, **5,** 293-309.

Kazdin, A. E. and Bootzin, R. R. The token economy: an evaluative review. *Journal of Applied Behavior Analysis,* 1972, **5,** 343-372.

McLaughlin, F. T. and Malaby, J. Reducing and measuring inappropriate verbalizations in a token classroom. *Journal of Applied Behavior Analysis,* 1972, **5,** 329-333.

O'Leary, K. D. and Drabman, R. Token reinforcement programs in the classroom: A review. *Psychological Bulletin,* 1971, **75,** 379-398.

Paul, G. L. Chronic mental patient: current status—future directions. *Psychological Bulletin,* 1969, **71,** 81-94.

Phillips, E. L. Achievement Place: token reinforcement procedures in a home-style rehabilitation setting for "predelinquent" boys. *Journal of Applied Behavior Analysis,* 1968, **3,** 213-223.

Weisberg, P., Lieberman, C., and Winter, K. Reduction of facial gestures through loss of token reinforcers. *Psychological Reports,* 1970, **26,** 227-230.

Winkler, R. C. Management of chronic psychiatric patients by a token reinforcement system. *Journal of Applied Behavior Analysis,* 1970, **3,** 47-54.