SELF-EVALUATION BY ADOLESCENTS IN A PSYCHIATRIC HOSPITAL SCHOOL TOKEN PROGRAM[1]

DAVID A. SANTOGROSSI, K. DANIEL O'LEARY,[2]
RAYMOND G. ROMANCZYK,[3] AND KENNETH F. KAUFMAN

STATE UNIVERSITY OF NEW YORK AT STONY BROOK AND
SAGAMORE CHILDREN'S CENTER, MELVILLE, NEW YORK

Nine adolescent boys with a history of high rates of disruptive classroom behavior were selected from a psychiatric hospital school and placed in a remedial reading class after school in which various factors in a token reinforcement program involving self-evaluation were investigated. The effects of self-evaluation, in the form of a rating the students gave themselves about the appropriateness of their classroom behavior, were first assessed. While the students' ratings of their own behavior correlated highly with the teacher's ratings and evaluations made by independent observers, the self-evaluations did not lead to a reduction in disruptive behavior. A token reinforcement program, in which the teacher rated the students' level of appropriate behavior and in which the students traded earned rating points for prizes, clearly led to a reduction of disruptive behavior. When the students were given the opportunity to evaluate their own behavior and to receive rewards in exchange for the evaluation, they returned to their former rates of disruptive behavior.

Withdrawal of classroom token programs typically occurs by the sudden cessation of token and backup reinforcer availability, or by gradual thinning of such reinforcers, *i.e.,* increasing the amount of desirable behavior required for each unit of reinforcement. An alternative procedure that might be utilized in the withdrawal of token programs would be to transfer regulation of token programs from teacher to student control. This procedure would remove some responsibility for classroom discipline

from the teacher and allow him to devote his energies more fully to creative instruction. Silberman (1970) offered anecdotal evidence that children allowed to participate in the design and maintenance of their learning environment are happier and more productive students. Lovitt and Curtiss (1969) also found that a child allowed to specify his own contingencies worked harder than when similar contingencies were imposed upon him. In addition to being effective (Drabman, *unpublished*), student-regulated token programs may also be more philosophically acceptable to those for whom the idea of control by conspicuous external agents is offensive.

Self-monitoring of target behaviors has been shown to have some effectiveness in modifying behaviors such as multiple tics (Thomas, Abrams, and Johnson, 1971), classroom studying and talking out (Broden, Hall, and Mitts, 1971), and smoking (McFall, 1970). Recent studies have claimed to demonstrate the response maintenance effects of instructing subjects to reinforce their own performance on behaviors such as concept identification (Kanfer, Bradley,

[2]Reprints may be obtained from K. Daniel O'Leary at the Department of Psychology, State University of New York, Stony Brook, New York, 11790.

[3]Present address: Psychological Clinic, Rutgers-The State University, New Brunswick, New Jersey.

and Marston, 1962), dart-throwing (Marston, 1967), wheel-cranking (Bandura and Perloff, 1967) and verbal discrimination (Kanfer and Duerfeldt, 1967). Subjects were instructed to self-reinforce either by signalling correct responses or by selecting reinforcers from an available supply. Such self-reinforcement procedures typically maintained behaviors for as long a time as experimenter-administered reinforcement methods. However, almost all of the self-reinforcement studies have been conducted with normal subjects and the behavior of the subjects was assessed for only brief periods.

In a study by Glynn (1970), normal subjects were awarded tokens on the basis of their report of academic performance. For one group, amounts were determined by the subjects themselves, while for another group, tokens were self-administered, but at a rate determined by the experimenter. Both groups proved equally superior to a chance-determined token group and a no-token control group in maintaining and improving academic achievement as reported by the students. These group rankings maintained when all three experimental groups adopted the self-determined tokens procedure. However, these group differences were quite small and their significance is questionable due to the lack of appropriate initial group matching. In addition, no independent measures of student performance were reported to document academic improvement.

Several critical reviews of behavior modification have suggested that various self-evaluation or self-reinforcement procedures be incorporated into token reinforcement programs (Kazdin and Bootzin, 1972; O'Leary and Drabman, 1971), but only one published study exists in which severely deviant youngsters were given the opportunity to evaluate their own behavior in a token program (Kaufman and O'Leary, 1972). The students behaved extremely well during the self-evaluation period, which lasted for seven days, but an even longer period of self-evaluation is needed before one would suggest that self-evaluation be introduced as a

standard procedure in all token programs. One might surmise that at least some children would greatly over-evaluate their behavior in order to receive rewards associated with high evaluations.

In the Kaufman and O'Leary study, unruly, emotionally disturbed adolescents in a psychiatric hospital school were placed in token reinforcement programs in two remedial after-school classes. After the students had been in the token programs in which the teacher had evaluated the students' behavior and the students had exhibited very low rates of disruptive behavior for 25 days, the students were instructed to evaluate their own classroom behavior. That is, they were to evaluate how well they had adhered to the classroom rules. In the earlier token reinforcement period, the teacher had given students ratings exchangeable for prizes; in the self-evaluation phase, the students rated themselves publicly and received prizes commensurate with the ratings they had given themselves. In both classes, disruptive behavior remained at the previous low levels for the remainder of the program, six days in one group, seven in the other. It should be emphasized that students could have been disruptive yet evaluated their behavior as "very good" (*i.e.,* given themselves the highest rating) and still have received the highest valued prizes. Considering the initial high levels of disruptive behavior and the diagnostic classification of these children (behavior disorder or childhood schizophrenia), the maintenance of their good behavior for as long a period as seven days was somewhat surprising.

The first aim of the present study was to determine the effects of self-evaluation *per se* on disruptive behavior. Secondly, an effort was made to determine the duration of maintenance of appropriate behavior following transfer of control of point determination in a token program from teacher to students. Finally, the extent to which pupil evaluation was related to the evaluation of independent observers and teacher ratings was examined.

METHOD

Experimental Situation

An after-school remedial reading class in a children's psychiatric hospital was established specifically for this experiment. The class met for 30 min, five days a week, for a total of 51 days extending over nearly three months. Each session was divided into two 15-min periods. SRA reading laboratories (Parker, 1964) were employed as instructional material. The materials include reading selections followed by comprehension questions and exercises. Each pupil was able to begin at his own reading level and to proceed at his own pace. Pupils scored and recorded their own progress. The teacher provided individual assistance when requested.

The following five rules were in effect and posted at the front of the classroom throughout the experiment: no talking, face front of room, raise hands to speak, work hard, and work continuously. Within-subject comparisons were made of behavior changes over seven experimental conditions: (1) baseline, (2) self-evaluation alone, (3) teacher-determined points with back-up reinforcement I, (4) self-determined points with back-ups I, (5) matching of self and teacher ratings reinforced, (6) teacher-determined points with back-ups II, and (7) self-determined points with back-ups II.

Subjects

Nine adolescent boys reported by their regular classroom teachers to be deficient in reading skills and displaying high levels of disruptive classroom behavior participated in an after-school reading class. They were all residents in the psychiatric hospital for emotionally disturbed children where the study was conducted. Some characteristics of the subject sample determined from routine hospital records, are summarized in Table I.

Teacher

Mr. H., who had taught at the hospital school for 1 yr, instructed the reading class daily after regular classes. For his services, he received three graduate credits in psychology and a stipend of $300.00.

Table 1

Description of Pupils

| Subject | Age | | WISC | Psychiatric |
	yr.	mo.	Full Scale	Diagnosis
S1	12,11		94	Schizophrenia, childhood type
S2	13,11		124	Unsocialized aggressive reaction of childhood; Nonpsychotic OBS* w/epilepsy
S3	14,2		91	Personality disorder, explosive; Nonpsychotic OBS w/other unspecified physical conditions
S4	12,4		76	Schizophrenia, childhood type; Nonpsychotic OBS w/other unspecified physical conditions
S5	13,9		96	Nonpsychotic OBS w/other unspecified physical conditions
S6	15,1		92	Adjustment reaction of adolescence
S7	12,11		96	Passive aggressive personality; Nonpsychotic OBS w/other unspecified physical conditions
S8	14,0		87	Adjustment reaction of adolescence
S9	12,0		87	Schizophrenia, childhood type
Range	12,0-15,1		76-124	
Mean	13,5		96	
Median	13,9		92	

*OBS: Organic Brain Syndrome

Observation

Every pupil was observed during each class session throughout the study. Observations were made by undergraduate students. A total of 15 observers participated in pupil observation with a maximum of five regular observers and a reliability checker present on any one day. Five observers were always assigned to the nine students. Each of the five observers recorded the behavior of five pupils for the first 15-min period. Since there were five observers to monitor the remaining four pupils, during the second 15-min period two observers would watch the same pupil at the same time. The reliability checker also observed the same pupil, permitting him to check the reliabilities of two observers simultaneously. Thus, a total of three reliability checks could be obtained each session.

Each pupil was observed in random order for at least 15 min per class using the behavior codes and the method described by O'Leary, Kaufman, Kass, and Drabman (1970). Observations were made on a 20-sec observe, 10-sec record basis. The nine categories of disruptive behavior were:

1. Out-of-chair: movement of the child from his chair when not permitted or requested by teacher. No part of the child's body is to be touching the chair.
2. Modified out-of-chair: movement of the child from his chair with some part of the body still touching the chair (exclude sitting on feet).
3. Touching others' property: child comes into contact with another's property without permission to do so. Includes grabbing, rearranging, destroying the property of another, and touching the desk of another.
4. Vocalization: any unpermitted audible behavior emanating from the mouth.
5. Playing: child uses his hands to play with his own or community property so that such behavior is incompatible with learning.
6. Orienting: the turning or orienting re-

sponse is not rated unless the child is seated and the turn must be more than 90 degrees to either side, up or down, using the desk as a reference point.
7. Noise: child creating any audible noise other than vocalization without permission.
8. Aggression: child makes movement toward another person to come into contact with him (exclude brushing against another).
9. Time off task: child does not do assigned work for entire 20-sec interval. For example, child does not write or read when so assigned.

As many as nine categories of behavior could be recorded in any 20-sec interval. Only one instance of any category of disruptive behavior could be recorded in each 20-sec interval and the category was scored if the behavior was observed during any portion of the interval. The daily level of disruptive behavior was calculated by dividing the total number of disruptive behavior categories recorded by the total number of intervals observed.

The final condition was carried out after most observers were no longer available due to the end of the university semester. As a result, the data of the last four study days reflect a time-sampling procedure. Pupils were observed for four 30-sec intervals (a total of 2 min), each in a random sequence. Two of these 2-min periods were included during each class session for each pupil. Correlation of 20 random 2-min intervals from earlier conditions with their respective 15-min observation periods was 0.94, indicating that shorter periods were adequately representative of the longer ones.

Reliability of Observations

Reliability of observer agreement was scored when both the observer and an independent reliability checker recorded the same category of disruptive behavior during the same 20-sec observation interval. Disagreements were scored when one observer recorded a behavior category

and the other did not. Reliability was computed by dividing the number of observer agreements by the total number of agreements plus disagreements. Each observer was checked for reliability at least once during each experimental condition (except for the Matching Phase, which lasted only one day).

A total of 127 reliability checks were made. Observer reliability for the measure of disruptive behavior throughout the study ranged from 0.60 to 1.00 and averaged 0.79. These figures are somewhat lower than those typically reported, but few, if any, previous studies have included the precaution to avoid informing each observer when his reliability was being checked.

Reliability is likely to be higher when observers are aware they are being checked (Romanczyk, Kent, Diament, and O'Leary, 1973). In efforts to avoid this pitfall, the reliability checker never disclosed in advance which observer was being monitored. Presumably, observers who believe they *may* be monitored will remain more alert.

PROCEDURE

Phase 1: Baseline

During the first nine class days, behavioral observations were made but no experimental manipulations were effected. In this way, base measures were provided against which subsequent changes could be measured. Subjects were reminded of the rules at the beginning of each 15-min time period but no methods beyond verbal praise, occasional reprimands, and ignoring of disruptive behavior were employed to promote compliance with classroom rules.

Phase 2: Self-Evaluation

To determine the possible effects on disruptive behavior of self-evaluation alone, this second condition, lasting six days, required subjects to rate their own behavior with no back-up reinforcers provided for exchange for points awarded. Subjects were instructed by the teacher

that the ability to evaluate honestly and objectively their own behavior was a valuable skill, necessary for the development of independence, maturity, and responsibility.[4]

At the end of each 15-min period, the teacher surreptitiously recorded a rating for each subject's behavior during that period. This procedure allowed comparison of teacher- and self-ratings with behavior rates recorded by observers. Each subject was then, in turn, required to announce publicly the rating he would award himself, because it was felt that public evaluations would diminish gross over-evaluations. The degree of adherence to the rules was rated by both teacher and pupils on a scale from zero to two for each of the five rules, with zero representing little or no cooperation, one representing partial cooperation, and two representing complete cooperation. Each subject could therefore receive as many as 10 points per period for a possible total of 20 per daily class session. No further consequence of point accumulation was provided. This phase lasted six days.

Phase 3: Teacher-Determined Points with Back-ups I

During the nine days of this condition, the teacher publicly awarded each pupil points for following the five rules, based on the same 0 to 2 scale, at the end of each 15-min period. No ratings were asked from the subjects. Points obtained were exchangeable at a token store for snacks, fruits, and inexpensive prizes which ranged in price from 2 to 150 points (approximately 2¢ to $1.50). Token exchange occurred immediately after class and was supervised by the teacher. It was predicted that, as in an earlier study with a similar token reinforcement program and similar subjects (Kaufman and O'Leary, 1972), disruptive behavior would be reduced to a low frequency during this phase.

---

[4]Copies of specific instructions given for each condition may be obtained from the authors.

Phase 4: Self-Determined Points with Back-ups I

An attempt was made to replicate the successful transfer from the teacher-determined to a student-determined token program achieved by Kaufman and O'Leary (1972) by continuing to provide back-up reinforcers in exchange for points, the points being self-awarded. Subjects applied their own public ratings as in the self-evaluation (Phase 2) condition, but in this phase the points were exchangeable for back-up reinforcers, as was the case during the teacher-run program (Phase 3).

Phase 5: Matching

If subjects resumed high rates of disruptive behaviors but awarded themselves high ratings in order to gain back-up reinforcers more readily, the contingencies would then be so arranged that lying would be encouraged. To promote veridical self-determination of points deserved, each subject was awarded three bonus points if his self-rating matched the teacher's rating within one point. Three points would be subtracted from the teacher rating if the subject over- or under-estimated his rating beyond this margin. For example, if a pupil rated himself at eight and the teacher awarded seven, the pupil would net 10 points, a combination of the teacher determined points plus three bonus points. However, if he self-rated 10, while the teacher rated eight, the student would lose three points from the teacher rating, for a total of 5. It was explained that this feedback would help the pupils become more accurate in the important skill of self-evaluation.

Phase 6: Teacher-Determined Points with Back-ups II

The manipulations involved in this condition were identical to those of Phase 3.

Phase 7: Self-Determined Points with Back-ups II

The conditions in effect during Phase 4 were reinstated during this phase.

RESULTS

Daily averages of group disruptive behavior throughout all conditions are presented graphically in Figure 1. Baseline data indicate that the subject group was initially extremely disruptive, averaging over 1.5 disruptive behaviors per 20-sec interval. Institution of the self-evaluation condition did not substantially decrease the rate of inappropriate behaviors. The overall average of this phase was 1.4 disruptive behaviors per interval.

A dramatic decrease in inappropriate behavior did occur immediately when a conventional token program was established, *i.e.*, the teacher administered points exchangeable for prizes. The average rate of disruptive behavior was 0.3 during this period.

The low level of disruptive behavior maintained for the first four days of the self-determined token program but resumed baseline level on the fifth day. Self-evaluation of their behavior by the subjects did not reflect this change. On the fifth day of this phase, subjects continued to award themselves high ratings as they had in the first four days, despite the fact that they were now engaging in a high degree of disruptiveness. Because of the arrangement of contingencies, it was possible that subjects could derive the highest prizes by giving themselves high evaluations even if they displayed high rates of disruptive behavior. Thus, subjects were, in effect, receiving reinforcement for lying at the end of the first self-determined point phase.

The subsequent condition, matching, was an abortive attempt to remedy this situation. Upon hearing the instructions that they were to attempt to match the teacher's evaluation within one point, the subjects protested strenuously and loudly, claiming they thought the task too difficult, and they became completely unruly, refusing to cooperate. Their anger was intensified at the end of the first 15-min period of the class session when some of them were penalized for failing to match. The level of the disruptive be-
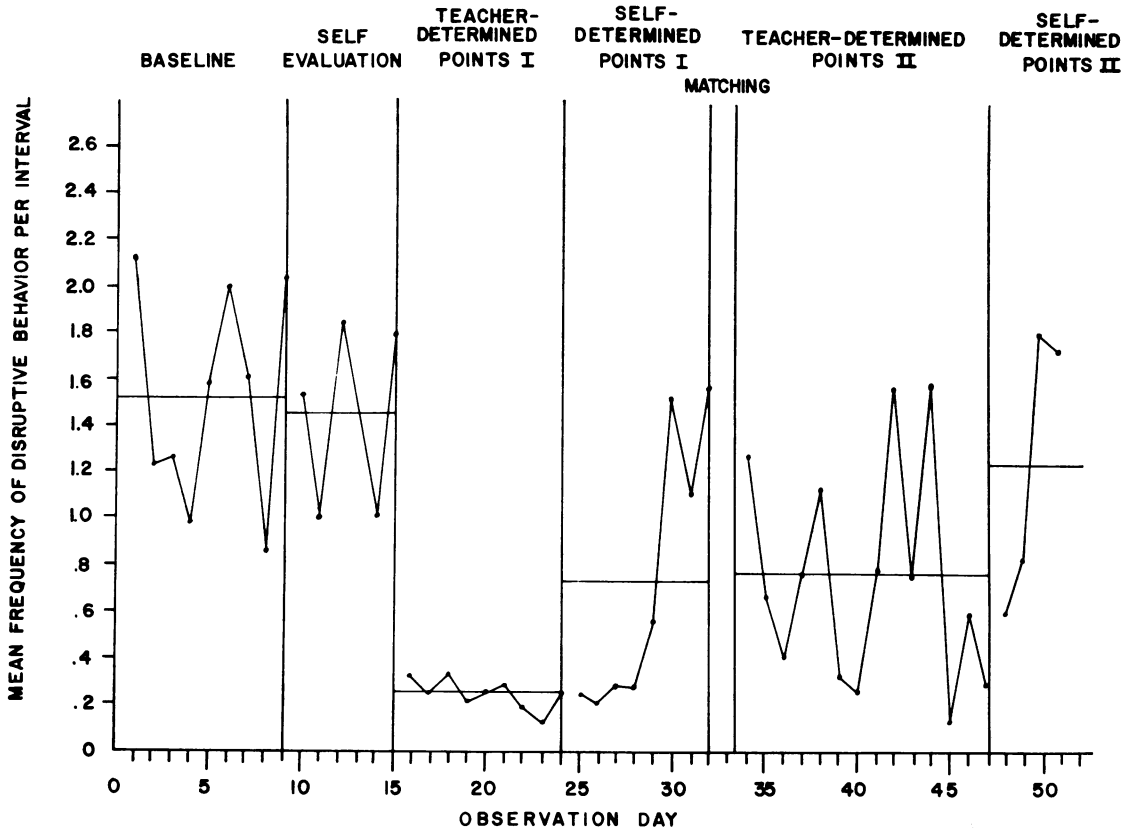
Fig. 1. Disruptive behavior of reading class as a function of experimental conditions.

havior for this one-day condition was 2.4 per interval, the highest level reached during the entire study. Consequently, this condition was abruptly terminated.

Subsequent reapplication of the standard token program situation seemed also to have been undermined by the matching phase manipulation. Subjects claimed they no longer valued the prizes available to them in the token store. On several days, some subjects decided they would not work at all, and these days show up as peaks in the Figure I graph of the Phase 6 condition. The mean level of disruptive behavior for this phase was 0.76 which, while still substantially below baseline level, is not as low as the original teacher-run token phase (Phase 3) rate.

When again given the opporunity to determine their own amounts of points in the final condition as in Phase 4, results similar to those

of Phase 4 were found. Subjects again quickly, though not immediately, resumed high rates of disruptive behavior. At the same time they also tended to overrate their levels of appropriate behavior.

Graphs for individual subjects generally show the trends the group data indicate, so that nine replications of effects noted are provided (see Figure 2). Nearly all subjects evidenced low disruptive behavior rates during initial days of of self-evaluation with points phases followed by an abrupt rise in inappropriate behavior. Subject 7 actually increased his frequency of disruptive behavior during the initial self-evaluation stage (without back-ups). Subject 9 did not participate in the rebellious reaction to the Matching Phase. Subjects 3 and 6 withdrew from the experiment during Phases 3 and 4, respectively, as a result of routine transfer to another hospital.
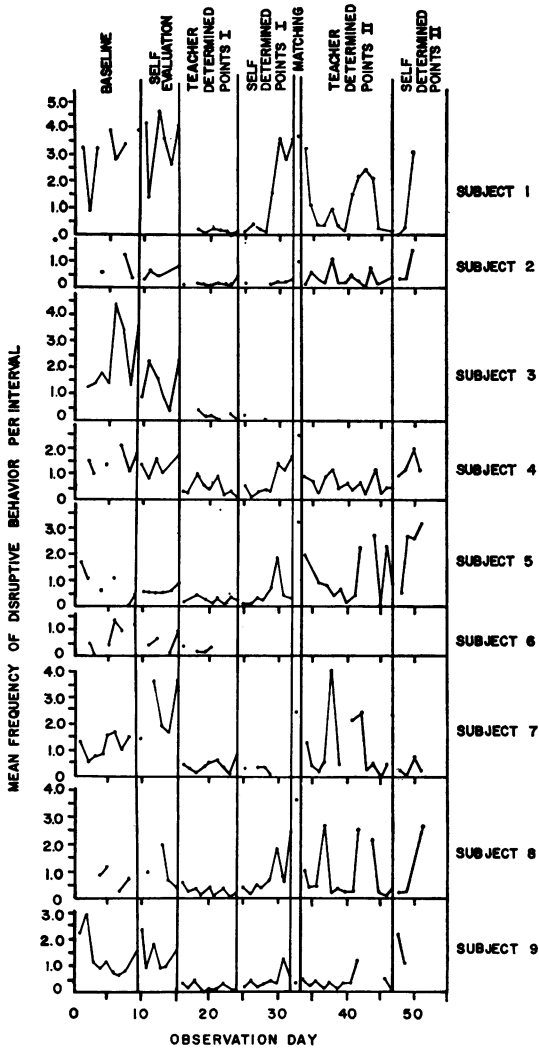
Fig. 2. Disruptive behavior of individual subjects as a function of experimental conditions.

The Pearson product-moment correlation between teacher ratings and observer measures of behavior frequencies was -0.81 during Phase 2 (Self-evaluation) and -0.83 during Phases 4 and 7 (Self-determined points I and II). During Phase 2, subject self-ratings correlated -0.70 with observer measures and $+0.95$ with teacher ratings. These correlations were -0.12 and $+0.52$ respectively, during Phases 4 and 7. Negative correlations occur because high teacher or student ratings would correspond to low frequencies of disruptive behavior. Figure 3 presents in graphical form class mean ratings as

awarded by the teacher and the students themselves during these conditions. The diagram indicates that teacher-and-student-rating correlations were low in the latter two conditions as a result of consistent over-rating by the students as their disruptive behavior increased. In general, the group means reflect individual data with two exceptions: Pupil 7 continued to rate accurately throughout both self-determined points conditions and Student 2 did likewise until the final day, when he underrated his own behavior.

DISCUSSION

Comparison of data from baseline and self-evaluation conditions reveal that self-evaluation alone was ineffective in reducing disruptive behavior significantly below baseline level. The abrupt drop in disruption that occurred with the introduction of the teacher-run token program was maintained for several days into the self-determined token condition, a finding similar to that of Kaufman and O'Leary (1972). However, rates of disruptive behavior in the present study resumed a high level after a short while. It should be noted, however, that in the Kaufman and O'Leary (1972) study, the children had been in a token program in which the teacher evaluated student behavior for 25 days, whereas in the present study, the students were in a similar token program for only nine days before being given responsibility for evaluating their own behavior and receiving prizes commensurate with their self-evaluations. It seems clear from this study that the students knew in the second self-determined point phase (Phase 7) that they could misbehave yet receive high-valued prizes by giving themselves high self-evaluations. However, as in the Kaufman and O'Leary study, as well as in Phase 4 of the present study, they did not immediately resume their former levels of disruptive behavior. Since Kaufman and O'Leary were forced to terminate their classes, there is no way of determining how long the low rates of disruptive behavior might have persisted. It is
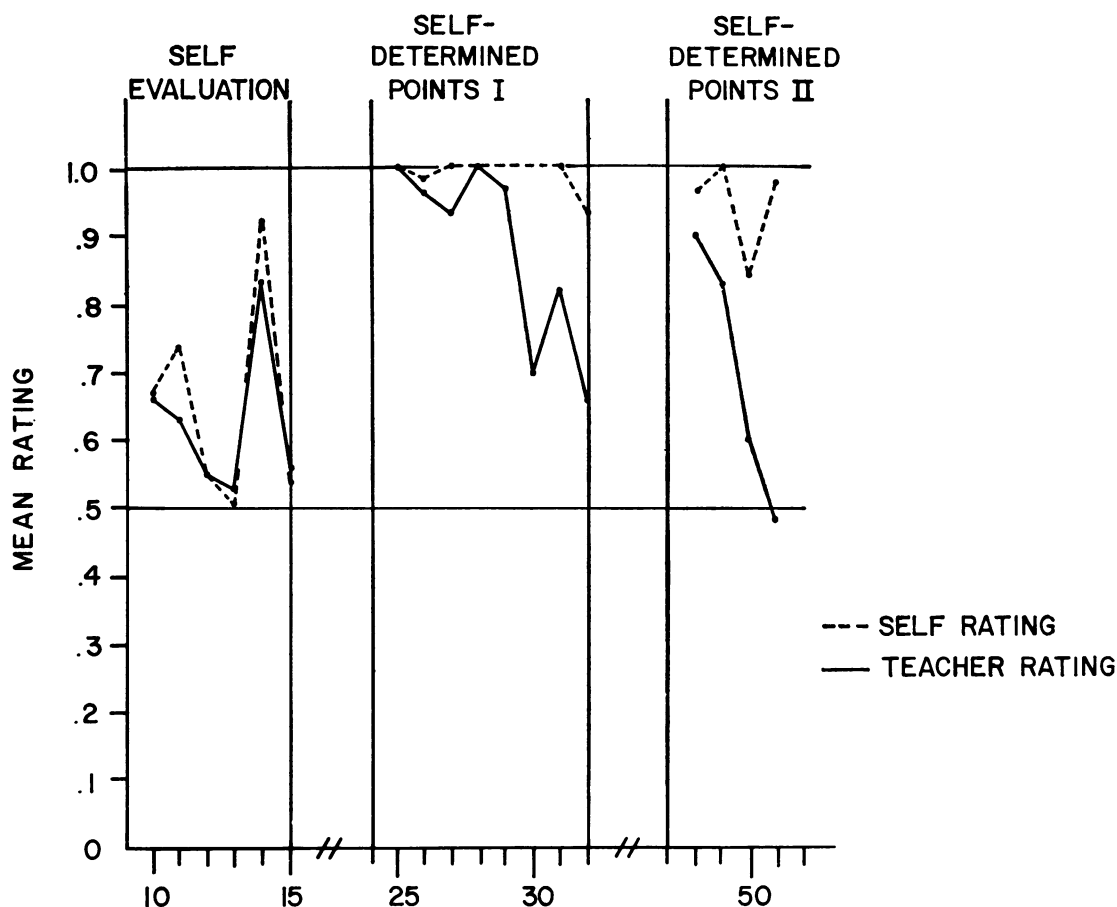
Fig. 3. Mean teacher- and self-ratings of student behavior during three experimental conditions.

possible that in the Kaufman and O'Leary token program (as they point out), academic skills had been improved to such a degree as to be incompatible with increased disruptive behavior on the part of the pupils (in both classes in their study significant gains were demonstrated in both amount of work completed and in reading ability). An alternative explanation may be that during the teacher-determined point phase a social milieu may evolve that encourages appropriate classroom behavior and discourages inappropriate behavior. It might be interesting to investigate whether the longer the token program in which the teacher evaluates behavior, the longer it will take for this milieu to dissolve.

The increases in disruptive behavior obtained in the first self-determined point phase seem to

be largely a consequence of the fact that two subjects (S1 and S5) quickly recognized that the system could be beaten, and on Days 5 and 6 of the first self-determined point phase, they were quite effective in employing social pressure to provoke other subjects into defiance. By declaring themselves "on strike" and denouncing as "fools" those who did not take advantage of the situation, these pupils appeared able to instigate much disruptive behavior. The second time students were given control of token determination, these two pupils in the second and third day of the condition again urged others not to cooperate.

The matching condition was undertaken in an effort to end lying and cheating by making verdical self-ratings more heavily reinforced than

inflated ones. Pupils in the present experiment rebelled against this manipulation, saying it would be too difficult. They claimed the teacher's ratings were inaccurate, having frequently disagreed with his assessments when he administered points during the teacher-run segments of the study. As noted earlier, however, teacher ratings and frequencies recorded by the independent observers correlated highly over all phases in which student evaluations were obtained. The close correspondence of teacher ratings and observed frequencies can also be noted by comparing Figures 1 and 3. These findings indicate that requiring students to match teacher ratings was appropriate in the sense that the teacher's rating was accurately related to standard observational recordings. Furthermore, during the initial self-evaluation phase, when no contingencies were applied to behavior or its rating, the overall correlation between teacher and student ratings was 0.95, indicating that the students were indeed able to match teacher ratings under some circumstances. Only when reinforcement was made contingent on high ratings did pupils consistently rate themselves highly despite their level of disruptive behavior (see Figure 3). Possibly the rates of disruptive behavior would have declined if the students had been kept in the matching phase for several days. Clearly, a value judgment rather than a scientific one was involved in the decision to terminate the matching phase in which there was a marked increase in disruptive behavior. The judgment was prompted by a concern that the school personnel would be antagonized by a continuation of the matching condition.

The failure of the matching condition was disappointing. Its failure may have been prevented by making variations in the procedures that shaped the children to match teacher rating correctly. Several possible variations could have been: (a) allowing students and teacher to rate simultaneously for a few days with the teacher giving students feedback about their accuracy in matching and giving reasons for his rating; (b) providing bonus points for pro-gressively better matching using initially lenient criteria for matching; (c) elimination or lessening of penalties for less inaccurate matches.

When continual monitoring and matching of pupil behavior and ratings are required of the teacher, no gains are accrued over more conventional teacher-run programs. In the present study, a matching system had been planned to be modified so the teacher would have monitored (rated) all his students on random days, or random students within the class each day and not announced that such monitoring had taken place until the end of class. At that time, veridical self-ratings would have been lavishly reinforced. Presumably, the uncertainty of the contingencies would have encouraged honest ratings. Such an arrangement might have provided the logistical and motivational advantages of a self-determined program while allowing teacher monitoring to prevent abuse.

As mentioned in the introduction, some investigators working with young normal children have found that behavior will be maintained as well under self-reinforcement as under experimenter-determined reinforcement conditions (Bandura and Perloff, 1967). That is, even when children are given the opportunity to set their own standards and give themselves tokens for certain behaviors, they display as many responses as when an experimenter sets the standards and gives them tokens for equivalent behaviors. However, one must be careful in generalizing from studies that are based on data collected during brief periods (i.e., single sessions or observations based only on several days). Care should be also be taken in generalizing from studies with normal children to other populations. Despite the fact that Kaufman and O'Leary found low rates of disruptive behavior in a self-evaluation condition of a token program with adolescent boys with behavior problems, the present study should alert behavior modifiers that caution is clearly indicated when one wishes to implement self-evaluation or self-reinforcement procedures with clinical populations.

REFERENCES

Bandura, A. and Perloff, B. Relative efficacy of self-monitored and externally imposed reinforcement systems. *Journal of Personality and Social Psychology,* 1967, **7,** 111-116.

Broden, M., Hall, R. V., and Mitts, B. The effect of self-recording on the classroom behavior of two eighth-grade students. *Journal of Applied Behavior Analysis,* 1971, **4,** 191-199.

Drabman, R. *A comparison of student controlled and teacher controlled token programs.* Unpublished doctoral dissertation. State University of New York at Stony Brook, 1972.

Glynn, E. L. Classroom applications of self-determined reinforcement. *Journal of Applied Behavior Analysis,* 1970, **3,** 123-132.

Kanfer, F. H., Bradley, M., and Marston, A. R. Self-reinforcement as a function of degree of learning. *Psychological Reports,* 1962, **10,** 885-886.

Kanfer, F. H. and Duerfeldt, P. H. D. Effects of pretraining on self-evaluation and self-reinforcement. *Journal of Personality and Social Psychology,* 1967, **7,** 164-168.

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.

Lovitt, T. C. and Curtiss, K. Academic response rate as a function of teacher- and self-imposed contingencies. *Journal of Applied Behavior Analysis,* 1969, **2,** 49-53.

Marston, A. R. Self-reinforced and external reinforcement in visual-motor learning. *Journal of Experimental Psychology,* 1967, **74,** 93-98.

McFall, R. M. Effects of self-monitoring on normal smoking behavior. *Journal of Consulting and Clinical Psychology,* 1970, **35,** 135-142.

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

O'Leary, K. D., Kaufman, K. F., Kass, R. E., and Drabman, R. S. The effects of loud and soft reprimands on the behavior of disruptive students. *Exceptional Children,* 1970, **37,** 145-155.

Parker, D. H. *Teachers handbook: SRA Reading Laboratory.* Chicago: Science Research Associates, 1964.

Romanczyk, R. G., Kent, R. N., Diament, C., and O'Leary, K. D. Measuring the reliability of observational data: a reactive process. *Journal of Applied Behavior Analysis,* 1973, **6,** 175-184.

Silberman, C. E. *Crisis in the classroom: The remaking of American education.* New York: Random House, 1970.

Thomas, E. J., Abrams, K. S., and Johnson, J. B. Self-monitoring and reciprocal inhibition in the modification of multiple tics of Gelles de la Tourette's syndrome. *Journal of Behavior Therapy and Experimental Psychiatry,* 1971, **2,** 159-171.