JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1970, 3, 123-132 NUMBER 2 (SUMMER 1970)

CLASSROOM APPLICATIONS OF SELF-DETERMINED REINFORCEMENT[1]

E. L. GLYNN

UNIVERSITY OF TORONTO

Self-determined, experimenter-determined, and chance-determined token reinforcement treatments were compared with a no-token treatment, in terms of effect on the learning of history and geography material in the classroom. Each treatment was assigned to one of four heterogeneous classes of Grade nine girls. An initial baseline period preceded the differential reinforcement period, and a token withdrawal period followed. Subsequently, the self-determined treatment was employed in all three token reinforcement classes, before a final baseline period occurred. Findings included a similarity of initial baseline performance for all classes, an equal superiority of self-determined and experimenter-determined treatments to chance-determined and no-token treatments, and significant improvement from initial baseline to final baseline for self-determined and experimenter-determined treatments, but not chance-determined and no-token treatments. Differential token reinforcement experience was found to influence subsequent rate of self-determined token reinforcement.

There is confusion in the thinking of educators on the use of extrinsic reinforcers in the control of children's classroom learning.

Clearly, educators do not object to extrinsic reinforcers *per se,* since grades, promotions, degrees, diplomas, and medals appear to enjoy the same widespread usage that Skinner noted in 1953. Moreover, it is difficult to imagine a classroom where teacher praise and reprimand are not used in an attempt to control children's behavior. Despite this widespread use of extrinsic reinforcement, there is objection to the employment of certain forms (such as candy and tokens) on the grounds that the student will become dependent on them and will be unable to perform without them. (Anderson, 1967.) Yet, surely, the same objection should hold against all forms of extrinsic reinforcement, including teacher praise and reprimand.

Perhaps an explanation for this confusion is that the operation of a token reinforcement system, more than the generally inconsistent operation of teacher praise and reprimand, emphasizes the extent to which children's behavior is under the control of an external agent. External control of behavior is distasteful to many educators who would agree with R. M. Gagné, that ". . . the student must be progressively weaned from dependence on the teacher or other agent external to himself." (Gagné 1965, p. 213.)

It is suggested that some of the confusion has resulted from equating extrinsic reinforcement with external control of behavior. The two terms are not interchangeable. Skinner (1953), in discussing self-control, suggests the individual may be capable of controlling his own behavior by means of dispensing his own reinforcement contingent upon making certain classes of responses. Various studies of self-reinforcement (Kanfer, Bradley, and Marston, 1962; Bandura and Kupers, 1964; and Bandura and Perloff, 1967), have permitted human subjects to take over the reinforcing function of the experimenter, by signalling correct responses, or rewarding themselves from a supply of tokens. Such self-administered reinforcing systems do seem to possess behavior maintenance capabilities, at least for simple responses—cranking a wheel (Bandura and Perloff, 1967), and visual discrimination (Kanfer and Duerfeldt, 1967).

The present study attempted to apply self-administered reinforcement procedures to classroom learning. If these procedures were to

[1]Reprints may be obtained from the author, Department of Education, University of Auckland, Box 2175, Auckland, New Zealand. Research reported in this paper was carried out in partial fulfillment of the Ph.D. degree at the Ontario Institute for Studies in Education (University of Toronto). The author is indebted to his chairman, Dr. S. B. K. Henderson for his valued support and encouragement.

prove effective, they may be more acceptable to educators because they suggest a way to wean children from dependence on an external agent, and at the same time, would permit the use of effective extrinsic reinforcers.

Three major purposes of the study were: (1) to compare the effectiveness of self-determined and experimenter-determined token reinforcement treatments in the classroom setting; (2) to examine the effects of token withdrawal following these treatments; and (3) to examine the effect of differential token reinforcement experience on the amount of subsequent self-determined token reinforcement. A distinction is made here between *determination* and *administration* of reinforcement. All token reinforcement in this study was self-administered, but the amount of reinforcement was experimenter-determined, chance-determined, or self-determined (within the limits imposed by the experimental procedure).

METHOD

Experimental Situation

The study was planned to require a minimum of accommodation on the part of teacher and children, since it was intended to test the practicability of token reinforcement within the regular classroom program. The study did not require the teacher to alter subject content or teaching methods. Four intact class groups were used, which meant that no changes in timetable were requested, and children were never removed from their usual class setting. The subject matter, history and geography, was taught to all four classes, in the same topical order, by the one teacher. The token reinforcement treatments were administered by one experimenter in all four classes.

Subjects

One hundred and twenty-eight ninth-grade girls, in four classes in a Toronto Separate School, served as subjects. Class size ranged from 30 to 34. Girls had been assigned to classes from an alphabetic list, which was divided into four sections. While not truly random, this procedure at least precluded deliberate stratification of classes according to ability.

None of the children presented any problem to the teacher with regard to disruptive behavior. The teacher considered all children "well-motivated" to learn, and interested in the subject matter. Having been present for a portion of the history and geography lessons of all four classes for the baseline period, the experimenter shared these opinions with the teacher.

Dependent Variables

1. Test performance. Working from a list of history and geography topics supplied by the teacher, the experimenter prepared 40 reading sheets, each of approximately 500 words. The class history and geography texts provided source material. Accompanying each reading sheet was a sheet of 20 five-option multiple-choice questions, based on the factual content of the reading sheets. Hence, the major dependent variable for each of the five phases of the study was the average number of test items correctly answered by each girl.

An attempt was made to match the readings and tests closely with the teacher's program. This was not always achieved because of the occasional need for the teacher to revise a topic before going on to the next, and because the experimenter was requested to produce several readings on New Zealand, at a time when the teacher had almost completed her coverage of the topic. The four classes received all readings and tests, in the same order, with each phase of the study containing approximately five history and five geography readings.

2. Performance-token ratios. The tokens were slips of paper, 2 by 1 in. (5 by 2.5 cm) bearing a star and the words "one credit". Tokens were exchanged for a variety of inexpensive prizes at the end of the first token phase, and at the end of the study. After each token reinforcement phase, a performance-token ratio was obtained for every girl. This ratio was formed by computing the total number of correct test items obtained in a particular phase, and dividing this number by the total number of tokens received in that phase.

Originally, it was planned to make the tokens exchangeable for the potentially reinforcing events available within the school program, and selected by children according to preference. Examples of such potentially reinforcing events are: time off a particular activity, library time, free time, homework exemptions, punishment exemptions, and the right to perform special duties. However, it was

discovered that control of such events was out of the hands of the teacher concerned. Secondary school teachers do not have the same freedom in manipulating timetables and reinforcing events as do elementary school teachers, where one teacher handles one class for almost all the academic program.

In view of this difficulty, it was decided to provide a series of prizes, which could be obtained by turning in credits. An opportunity was taken to make the prizes relevant to some of the history and geography material, by using numerous inexpensive New Zealand souvenir items. When turning in their tokens, the girls were allowed to select a prize from the many items displayed, according to their rank order in number of tokens earned. There was a sufficient range of prizes for even the last-ranked child to have some choice.

3. Inter-class communication. An attempt was made to measure the extent of inter-class communication that occurred during the study, since it was realized that performance of one class could also be influenced by the knowledge that other classes were receiving different treatments. A set of three open-ended questions was administered to all children at the end of the study, asking them whether the treatment given their own class differed from that given the other classes, and if so, to state how.

Procedure

1. First baseline phase (Baseline I). This was a two-week period that served to establish basal measures of test performance in each of the four classes, and to accustom the children to the presence of the experimenter and the testing procedures. No tokens were issued.

Each day, the children were given a passage to read for 3 min, immediately after which the passage was collected, and a further 3 min were allowed for the multiple-choice test. (In considering these short time limits, it should be noted that the material encountered on the reading sheets would also have been covered by current teacher lessons). When the 3-min test session was over, immediate feedback of results was given, by means of the experimenter reading out the letter code for correct answers. The children then counted the number of test items correct, and entered this on a slip of paper in individual envelopes supplied for the purpose. Finally, test sheets and envelopes were collected by the experimenter. Instructions stressed that information in the envelopes would not be made available to anyone other than the experimenter.

2. First token phase (Token I). This was the only period in which token reinforcement procedures differed across classes. The procedures employed were:

(a) Experimenter-determined token reinforcement. Under this treatment, children received tokens according to an explicit rate of one token per four correct answers. During the token reinforcement periods, five tokens were placed in each envelope each day. The children were instructed to calculate the number of tokens earned by dividing their test score by four. An arbitrary rule permitted the taking of an additional token for a fractional number. (Thus a child would take four tokens if the number earned were 3¼, 3½, or 3¾.)

(b) Self-determined token reinforcement treatment. Under this treatment, children were invited to: "decide how many tokens you think you should award yourself. You can decide on any number from zero to five." No rules or suggestions were made concerning bases for decision making. The use of envelopes was intended to minimize the effect of social cues from peers, and of modeling peer standards, both of which are known to influence the rate of self-reinforcement (Marston, 1964; Bandura and Whalen, 1966; McMains and Liebert, 1968).

(c) Chance-determined token reinforcement. This was, in effect, an incentive-control treatment. Throughout the first token phase, the total number of tokens received by this class was kept identical with that of the self-reinforced class. Each day, chance-reinforced children were randomly assigned a "partner" from among the self-reinforced children. Regardless of performance of the chance-reinforced child, she found in her envelope the number of tokens that her self-reinforced "partner" for the day had taken. As well as providing an incentive control treatment, this procedure enabled the examination of the effect of such inconsistent experience of amount of reinforcement on extent of subsequent self-reinforcement.

(d) No-token reinforcement treatment. No token reinforcement was given under this treatment. The procedure was exactly the same as during Baseline I.

3. First token withdrawal phase (Withdrawal I). During this phase, token reinforcement was withdrawn from experimenter-reinforced, self-reinforced, and chance-reinforced classes.

4. Second token phase (Token II). Tokens were reintroduced for the experimenter-reinforced, self-reinforced, and chance-reinforced classes, but all three classes were now permitted to operate the self-reinforced procedure. The major question asked was whether the children previously reinforced according to an externally imposed standard (experimenter-reinforced) would subsequently display a rate of self-reinforcement close to this standard.

5. Second baseline phase (Baseline II). During this phase, the baseline readings and tests were readministered, in order to compare increase in performance on re-learning among the four classes. Tokens were withdrawn for the first half (Withdrawal II), but included for the second half (Token III).

6. Review test. A review test was constructed of items from each of the tests administered during the first token phase, the token withdrawal, and the second token phase, in order to determine whether treatment effects were of a long-term nature. Since the test was administered after the repeat of baseline, the test-retest interval was 2 to 4 weeks for items from the second token phase (Sections C), 4 to 6 weeks for items from the token withdrawal phase (Section B), and 6 to 8 weeks for items from the first token phase (Section A). For the review test, seven tokens were provided in the envelopes of the experimenter-reinforced, self-reinforced, and chance-reinforced classes, and the self-reinforced procedure was applied in all three classes.

RESULTS

Daily test performance scores of all classes throughout the study are listed in Table 1. Also shown are the mean test performance scores for each phase of the study.

It was considered that the performance of the non-reinforced class provided the best available estimate of variations due to fluctuations in test difficulty. Accordingly, Fig. 1 was produced by depicting the daily performance of the three treatment classes, in terms of difference from the non-reinforced class, so that variation due to fluctuating test difficulty might be removed.

Baseline I

An analysis of variance performed on mean scores for Baseline I yielded a non-significant between-classes effect ($F\ 3,\ 116 = 1.65,\ p > 0.05$), and Hartley's test for homogeneity of variance yielded an F max. of 1.40 which is not significant. The four classes were thus regarded as being similar in performance during Baseline I.

Token I

An analysis of covariance was performed on Token I mean scores, using Baseline I mean scores as covariate. The between-classes effect was significant. ($F\ 3,\ 115 = 16.69,\ p < 0.001$). An analysis of variance for repeated-measures on test scores in every alternate session of the Token I phase yielded an insignificant classes-by-sessions interaction. ($F\ 12,\ 464 = 1.61,\ p > 0.05$). Hence, analyses of mean Token I scores do not conceal any useful information about differential performance of classes across sessions.

The significant between-classes effect noted above is evident in Fig. 1. Token I mean scores were adjusted for the effect of the covariate (Baseline I performance), by the method suggested by Winer (1962, p. 592). Comparisons were made among the adjusted means by the Newman-Keuls procedure. The experimenter-reinforced and self-reinforced classes did not differ from one another, nor did the non-reinforced and chance-reinforced classes differ from one another. However, both the experimenter-reinforced and the self-reinforced classes differed significantly from the non-reinforced and chance-reinforced classes. The self-determined reinforcement procedure was equally as effective as the externally determined one, in producing an increase in performance.

Withdrawal I

The analysis of covariance performed on Withdrawal I mean scores (using Baseline I mean scores as covariate), yielded a significant between-classes effect ($F\ 3,\ 115 = 3.24,\ p < 0.05$), though the effect was weaker than that of Token I. This was expected, since in Token I, performance was directly influenced by differences in treatment procedures. Again, a repeated-measures analysis of variance yielded a non-significant classes-by-sessions interaction

Table 1
Mean Performance Scores for all Sessions and Phases

Session		Class			
		Non-Reinforced	*Chance-Reinforced*	*Experimenter-Reinforced*	*Self-Reinforced*
Baseline I:	1	10.60*	9.57	9.88	9.66
	2	11.60*	10.54	11.13	11.47
	3	9.69*	8.80	8.70	8.35
	4	9.82	8.63	8.91	9.91*
	5	10.71*	10.19	9.47	8.70
	6	12.21*	10.87	9.81	10.69
	7	9.72*	6.97	8.39	9.25
	8	10.55*	9.32	9.81	9.70
	9	9.46*	6.55	7.44	7.25
	10	9.63	9.42	8.78	9.97*
	Mean:	10.38*	9.02	9.13	9.47
Token I:	11	7.44	7.14	8.33*	8.33*
	12	10.08	9.37	11.29*	10.15
	13	9.38	9.53	11.08	12.66*
	14	10.30	9.37	11.32	11.45*
	15	8.07	6.28	9.14	10.14*
	16	10.53	10.41	11.62	12.54*
	17	9.57	8.07	10.32*	10.07
	18	11.50	11.33	13.97*	13.27
	19	11.17	11.13	12.41	12.56*
	20	10.07	10.09	12.06*	11.61
	Mean:	9.71	9.13	11.12*	11.09
Withdrawal I:	21	11.25	9.20	11.61*	10.67
	22	10.53	8.75	10.52	11.10*
	23	12.83*	12.52	12.75	12.39
	24	13.31	13.79	13.63	14.48*
	25	11.29	10.86	10.94	12.73*
	26	8.44*	7.39	8.44*	7.40
	27	10.62*	9.46	9.74	10.23
	28	10.31*	8.87	10.13	8.80
	29	10.69	10.03	9.70	11.23*
	30	8.59*	6.83	8.55	8.32
	Mean:	10.44	9.72	10.53	10.87*
Token II:	31	11.12*	9.42	9.48	10.16
	32	11.97	10.55	11.84	12.28*
	33	9.89*	8.58	9.48	9.33
	34	11.47	10.42	11.63*	11.10
	35	9.17	8.10	9.52	9.87*
	36	11.50	9.46	11.28	13.22*
	37	8.36	7.52	8.76	9.13*
	38	9.63	8.83	11.03*	9.38
	39	9.83	10.34	11.10	12.16*
	40	12.30	11.45	12.65*	11.80
	Mean:	10.47	9.33	10.64*	10.59
Baseline II:					
(a) Withdrawal II:	41	10.90	11.42	12.42	12.70*
	42	13.64	12.55	13.84	14.09*
	43	10.36*	8.13	10.06	10.00
	44	8.36	9.97	10.91*	10.90
	45	11.57	10.40	11.25	11.07
	Mean:	10.74	10.49	11.79	11.87*
(b) Token III:	46	11.38	11.04	12.78	13.19*
	47	10.29	9.50	10.42*	10.10
	48	10.72	10.72	10.83	11.57*
	49	10.59*	8.93	9.37	9.83
	50	11.14	10.41	11.23	12.35*
	Mean:	10.48	9.78	10.89	10.96*

*Indicates highest scoring class.

(F 12, 464 = 1.17, $p > 0.05$) so that analyses of mean scores for Withdrawal I did not conceal information about differential performance of classes across sessions. The significant effect reported above suggests that there remained some effects of Token I treatments during

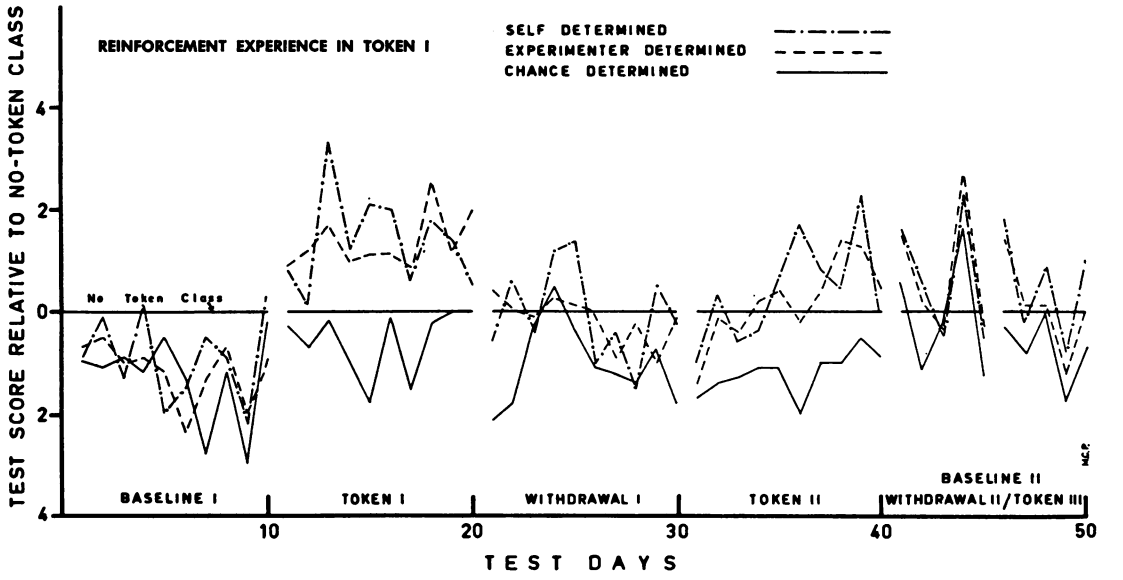

Fig. 1. Daily test performance of the three token classes relative to the no-token class.

Withdrawal I. This can be seen in Fig. 1. After the mean scores from Withdrawal I had been adjusted for the effect of the covariate, Newman-Keuls comparisons were made among them. It was found that the experimenter-reinforced and self-reinforced classes performed better than the non-reinforced class, but not better than the chance-reinforced class.

Token II

The analysis of covariance performed on Token II mean scores (using Baseline I mean scores as covariate) yielded a significant, though weak, between-classes effect ($F\ 3, 115 = 2.85$, $p < 0.05$), but none of the possible comparisons among adjusted means yielded significant differences at the 0.05 level by Newman-Keuls tests. However, it can be seen from Fig. 1 that the experimenter-reinforced and self-reinforced classes were superior to the non-reinforced class for the greater part of Token II, while the chance-reinforced class always remained inferior to the non-reinforced class, suggesting some differences in performance between classes.

A repeated-measure analysis of variance of Token II data yielded a significant classes-by-sessions interaction ($F\ 12, 464 = 2.47$, $p < 0.01$). Hence, Token II mean scores obscure differential performance of classes across sessions. Figure 1 shows that the experimenter-reinforced and self-reinforced classes displayed a more obvious improvement in performance relative to the non-reinforced class than did the chance-reinforced class.

Baseline II

As the readings and tests administered during Baseline II were the same as those for Baseline I, comparisons of performance gains over the two administrations were made across classes.

Highly significant phase effects were found, for both the Withdrawal II and Token III halves of the repeated-baseline tests, but these may be readily attributed to general practice effects. However, significant classes-by-phases interactions were also found: $F\ 3, 116 = 4.23$, $p < 0.01$ (for Withdrawal II), and $F\ 3, 116 = 2.80$, $p < 0.05$ (for Token III). Newman-Keuls comparison were carried out to examine these effects further. Table 2 shows that the inter-

Table 2
Newman-Keuls qr Values for Differences between Baseline I and Baseline II Mean Test Performance Scores.

Test Days	*Class*	*Baseline I*	*Baseline II*	*qr*
1-5	NR	10.49	10.74	0.84
and	YR	9.43	10.49	3.53
41-45	ER	9.55	11.79	7.42**
	SR	9.49	11.87	7.86**
6-10	NR	9.84	10.48	1.77
and	YR	8.58	9.78	3.32
46-50	ER	8.66	10.89	6.18**
	SR	8.81	10.96	5.96**

**$p < 0.01$

action effect is attributable to significant increases on Baseline II administration for the experimenter-reinforced and self-reinforced classes. This increase is not significant in the case of the non-reinforced and chance-reinforced classes.

Review Test

Analysis of variance performed on scores from each section of the Review Test yielded a significant between-classes effect ($F\ 3, 108 = 4.47$, $p < 0.01$) for Token I items (Section A). Newman-Keuls comparisons of class means revealed a pattern of significant results parallel to that of Token I performance. On Section A items, both the experimenter-reinforced and self-reinforced classes ($x = 9.89$ and 9.71) were slightly better than the non-reinforced and chance-reinforced classes ($x = 8.21$ and 8.00). There was no significant difference between the self-reinforced and experimenter-reinforced classes, nor between the non-reinforced and chance-reinforced classes. The performance increments resulting from differential reinforcement schedules evidently have some permanence. Sections B and C of the Review Test (Withdrawal I and Token II items) yielded no between-classes effects that reflected earlier treatments.

Performance-Token Ratios

Table 3 presents data concerning number of tokens taken, and performance-token ratios of the three token classes throughout the study.

Analysis of variance indicated no significant differences in number of tokens taken by the three classes during Token I ($F\ 2, 90 = 0.453$, $p > 0.25$). Hence, amount of token reinforcement can be regarded as similar across classes. Token I performance-token ratios were subjected to analysis by a median test for independent groups (Hays, 1962), since variances for the three groups departed widely from homogeniety (F max. $3, 30 = 9.05$, $p < 0.01$). The observed Chi-squared value for the median test was 35.1 ($p < 0.001$). Hence, despite similarity of amount of token reinforcement during the Token I phase, the performance-token ratio for the self-reinforced class was higher than those of the experimenter-reinforced and chance-reinforced classes, indicating that the self-reinforced class had "worked hardest" per token.

Similar results emerged in Token II. Again, no significant difference in amount of token reinforcement was found ($F\ 2, 90 = 2.37$, $p > 0.05$), but significant differences were found in performance-token ratios, ($F\ 2, 90 = 3.18$, $p < 0.05$), with that of the self-reinforced class being higher than those of the experimenter-reinforced or chance-reinforced classes.

In Token III, analysis of variance revealed that classes did differ in terms of amount of token reinforcement taken ($F\ 2, 90 = 4.19$, $p < 0.05$) with the self-reinforced class taking fewer tokens than the experimenter-reinforced class, though more than the chance-reinforced class. Yet, as Table 3 shows, the self-reinforced class again displayed the highest performance-token ratio.

The above pattern of results was also found in the data from the Review Test. There were significant differences both in amount of token reinforcement taken by the three classes ($F\ 2, 81 = 62.59$, $p < 0.001$), and in the performance-token ratios ($F\ 2, 81 = 9.20$, $p < 0.001$). It can be seen from Table 3 that the self-reinforced class again took fewer tokens than the experimenter-reinforced and chance-reinforced classes, and displayed the highest performance-token ratio on the Review Test.

Table 4 supplies information on the variability in performance-token ratios for all token reinforcement phases. Clearly, the ex-

Table 3

Number of Tokens Taken and Performance-Token Ratios in all Token Phases

	Chance-Reinforced Class		*Experimenter-Reinforced Class*		*Self-Reinforced Class*	
Phase	*Number*	*Ratio*	*Number*	*Ratio*	*Number*	*Ratio*
Token I	2.87	3.26	3.10	3.60	2.90	3.99
Token II	2.63	3.54	3.13	3.43	2.76	3.92
Token III	2.83	3.71	3.29	3.33	2.95	3.89
Review Test	3.82	7.77	5.82	5.77	3.57	9.50

Table 4
Standard Deviations of Performance-Token Ratios

Phase	Chance Reinforced Class	Experimenter-Reinforced Class	Self-Reinforced Class
Token I	0.98	0.43	1.31
Token II	1.01	0.59	0.63
Token III	1.12	0.62	0.76
Review Test	3.86	1.86	3.69

perimenter-reinforced class displayed the least variability throughout.

Communication Between Classes

On the open-ended questions concerning knowledge of the treatment of other classes, the maximum "information score" was 3.0. Mean scores for the four classes were: non-reinforced, 0.74; chance-reinforced, 0.73; experimenter-reinforced, 0.44; and self-reinforced, 0.27. Newman-Keuls comparisons showed the self-reinforced class mean to be significantly lower than those of the chance-reinforced and non-reinforced classes. This could indicate that reported performance differences might be confounded by other factors arising from amount of information about other classes. However, a detailed examination of responses to the open-ended questions suggested that the extent of any such confounding was not great. Children made surprisingly few statements about how procedures in any of the classes had differed from their own.

DISCUSSION

1. Effectiveness of Self-Determined Reinforcement

Restrictions must be placed on generalizing from the findings of this study, both in terms of the particular children involved, and in terms of the narrow range for self-determined reinforcement permitted by the procedure. It is an open question as to whether these results would be replicated with younger or underpriviledged children, without some modification of procedure—for example, providing a wider variety of more meaningful prizes. It is also an open question as to whether similar results would have been obtained with these children, had there been wider limits allowed on amount of reinforcement, and had there been no check by the experimenter on the amount of reinforcement taken. Furthermore, these results would appear more convincing, had there been a further non-reinforced control class at another school. This would have yielded direct information about the effect on the non-reinforced class, of knowledge of reinforcement contingencies in the other classes. More accurately than the self-report measures used in this study, it would determine whether the performance of the non-reinforced class during token phases, was, in fact, only a reflection of test difficulty, or whether it was confounded with adverse motivational effects arising from knowledge of other treatments.

Nevertheless, the study does suggest that the concept of self-determined reinforcement is both applicable and appropriate for studies of academic performance in the classroom. Self-determined reinforcement, within the above-mentioned limits, proved to be at least as equally effective as experimenter-determined reinforcement, in terms of improving academic performance. Children were able to control successfully the token reinforcement for their classroom learning, when both social cues and specific instructions about extent of reinforcement were minimized. It would seem that the notion of systematic social reinforcement as a "critical component" of an effective token system (Kuypers, Becker, and O'Leary, 1968) may need to be qualified.

It is clear that token reinforcement procedures were less effective in Token II and Token III than in Token I. Since tokens were exchanged for prizes for the first time at the end of Token I, and since identical sets of prizes were available at the end of Token III, it is thought that the tokens dropped much of their value as reinforcers. There is a need for future studies to ensure a sufficiently varied set of reinforcing events to back up the tokens. The particular prizes used in this study were nevertheless effective during Token I, possibly because of their novelty.

2. Performance of the Chance-Reinforced Treatment Class

This class performed at a level generally below that of the non-reinforced class throughout the study. The inconsistent experience of this class in terms of amount of reinforcement during the Token I phase, seems to have not only precluded performance increments during this phase, but also to have prevented subsequent self-determined reinforcement proce-

dures from having any incremental effect. This is certainly an indication that the ability to apply self-determined reinforcement is strongly influenced by the standards of externally determined reinforcement previously experienced. Hence, inconsistency of reinforcement can occur not only in terms of interpersing reinforcement with non-reinforcement as consequences of a given behavior, but also in terms of unpredictable amounts of reinforcement for a given behavior. These results suggest that parents and teachers, who function as major external reinforcing agents for children's behavior, should be aware that one consequence of maintaining such inconsistent standards of reinforcement may be impairment of the child's ability to apply self-determined reinforcement procedures effectively. If such an ability is considered as one component of self-control, as Marston and Kanfer (1963) suggest, then inconsistent experiences of amount of reinforcement would have a debilitating effect on the development of an individual's ability to control his own behavior.

3. Withdrawal of Tokens

Findings suggest that after token withdrawal, the four classes did not revert to the similarity of performance displayed during the baseline. Token reinforcement classes experimenter-reinforced and self-reinforced remained slightly superior to the non-reinforced class. There seems little evidence to justify the fear that children would become dependent upon token reinforcement so as to be unable to perform without it.

4. Performance-Token Ratios

Data on performance-token ratios provide further support that the operation of self-determined reinforcement is influenced by standards of externally determined reinforcement previously experienced. Table 3 shows that the experimenter-reinforced class adhered more closely to the performance-token ratio experienced during the Token I phase than did either the self-reinforced or chance-reinforced class. The experimenter-reinforced class had been supplied with an explicit ratio, whereas the self-reinforced and chance-reinforced classes had not. Yet, the self-reinforced and chance-reinforced classes moved towards a much higher performance-token ratio, especially on the Review Test. Table 4 indicates that the experimenter-reinforced class displayed the least variability in ratios throughout the study. This would be expected if members of this class were adhering to a common standard. The striking finding is that the children who had the greatest opportunity for leniency in taking tokens (self-reinforced class), actually imposed the strictest ratio on themselves.

The performance-token ratios observed in this study imply that an alternative to a teacher laying down explicit acceptable standards of performance for classroom learning, might be the provision of access to reinforcement on the basis of standards determined by individual children.

5. Applicability of Procedures

The token-reinforcement procedures employed proved to be well suited to classroom use. Tokens did not have to be paid out individually to each child (a saving of time and energy for the teacher). Handing out the envelopes took about 1 min each day, and children took about the same time to take their tokens and return the envelopes. Since envelopes contained a slip bearing daily performance scores, a continuous record was available showing performance and number of tokens taken. For experimental purposes, it can be noted that by including differential instructions in envelopes, several reinforcement procedures might be operated simultaneously.

REFERENCES

Anderson, R. C. Educational Psychology. *Annual Review of Psychology*, 1967, **18**, 129-164.

Bandura, A. and Kupers, C. J. Transmission of self-reinforcement through modeling. *Journal of Abnormal and Social Psychology*, 1964, **69**, 1-9.

Bandura, A. and Whalen, C. K. The influence of antecedent reinforcement and divergent modeling cues on patterns of self reward. *Journal of Personality and Social Psychology*, 1966, **3**, 373-382.

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

Gagné, R. M. *The conditions of learning*. New York: Holt, Rinehart & Winston, 1965.

Kanfer, F. H., Bradley, M. 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. Motivational properties of self-reinforcement. *Perceptual and Motor Skills*, 1967, **25**, 237-246.

Kuypers, D. S., Becker, W. C., and O'Leary, K. D. How to make a token system fail. *Exceptional Children,* October 1968, 101-117.

Marston, A. R. Variables affecting incidence of self-reinforcement. *Psychological Reports,* 1964, **14,** 879-884.

Marston, A. R. and Kanfer, F. H. Human reinforcement: experimenter and subject controlled. *Journal of Experimental Psychology,* 1963, **66,** 91-94.

McMains, M. J. and Liebert, R. M. Influence of discrepancies between successively modeled self-reward criteria on the adoption of a self-imposed standard. *Journal of Personality and Social Psychology,* 1968, **8,** 166-171.

Skinner, B. F. *Science and human behavior.* New York: Macmillan, 1953.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Received 13 October 1969.
(Revised 9 April 1970.)