*SOME EFFECTS OF TOKEN REWARDS ON SCHOOL
ACHIEVEMENT OF CHILDREN WITH DOWN'S SYNDROME¹*

A. J. DALTON, C. A. RUBINO, AND M. W. HISLOP

MENTAL RETARDATION CENTRE, TORONTO, CANADA

The effectiveness of a token economy system in producing improvement in the academic performance of children with Down's syndrome was tested. One group of seven children received token reinforcement for correct responses and showed significant improvement both in arithmetic and language. A second matched group of six children received only verbal praise for correct responses to the same instructional materials and failed to improve in arithmetic but showed significant gains in language. Re-test scores one year later revealed that the Token Group maintained its gains in both subjects whereas the language performance of the No-Token Group showed a significant decline.

Behavior modification research with children of normal intelligence (*e.g.*, Haring, Hayden, and Nolen, 1969; Miller and Schneider, 1970; Chadwick and Day, 1971) has amply demonstrated the effectiveness of token economy systems in producing improvement when successful performance on academic subjects is selected as the "target behavior" for modification. There are few comparable studies with retarded children (*e.g.*, Birnbrauer, Wolf, Kidder, and Tague, 1965). The purpose of the present study was to provide additional empirical data bearing on this problem.

Specifically, an attempt was made to test whether a group of children with Down's syndrome would make significantly greater gains in certain school subjects when such learning was supported by a token economy system than when the same material was taught to a second group of comparable children without the support of a

token economy. Both short- and long-term effects of the token economy were examined by undertaking re-tests of the same children 1 yr later. The possible generalization effects of the token economy were also assessed by daily ratings of each child's classroom conduct.

METHOD

Subjects

Table 1 provides the basic description of the children after their assignment to the Token Group or the No-Token Group. Assignment to either group was random with the constraints that overall, the I.Q. scores, age, and academic achievement scores at the start of the program would have similar distributions and means in the two groups. The subjects were 13 children with Down's syndrome living at home. The children, both male and female, ranged in age from 6 to 14 yr with an average age of 9 yr and nine months. The children obtained Leiter International Performance Scale I.Q. scores ranging from 30 to 64 points with an average I.Q. of 45. Children with Down's syndrome were selected because they provide a relatively homogeneous group of individuals. In addition, they suffer from a relatively common and readily identifiable condition that rarely involves serious physical handicaps necessitating drug treatments.

¹The authors are indebted to the following for their indispensable assistance in the collection and analysis of the data presented in this report: Leslie Cavanaugh, John Laporta, James Leonidas, Brian McKinnon, Marna Ramsden, Doreen Sears, Faith Slattery, Bryan Stanish, Tom Stuckey, and Linda-Mary Tatlock. The cooperation and assistance of Mrs. Jean Bruce, Principal, Beverley School, Toronto, is also gratefully acknowledged. Reprints may be obtained from A. J. Dalton, Mental Retardation Centre, 2 Surrey Place, Toronto 181, Ontario.

Table 1

Description characteristics of the experimental (Token) and control (No-Token) groups at the start.

Subject Number	Sex	Age		Leiter I.Q.	Conduct Scores
		Years	Months		Percent
TOKEN GROUP					
1	female	10	7	36	20.5
2	male	9	8	33	16.5
3	male	13	2	48	19.5
4	female	6	1	64	26.5
5	female	10	9	52	8.5
6	female	7	7	48	40.5
7	male	14	2	51	9.5
NO-TOKEN GROUP					
8	male	9	1	30	31.0
9	female	11	7	35	17.5
10	male	10	2	47	22.5
11	female	8	0	55	21.5
12	female	9	7	42	29.0
13	male	6	7	58	16.0

Setting

The study was conducted over an eight-week period during the children's summer vacation months. Each classroom was equipped with desks, a bookcase, and a blackboard. At irregular intervals throughout the study a tripod-mounted television camera was an added feature in each room. During such occasions, video-recording and -monitoring of teacher, child, and behavior rater activity were conducted from an adjoining hallway.

Measurements

1. *Conduct scores.* Conduct scores per cent for each child were obtained using a behavior rating scale described by Bijou, Peterson, Harris, Allen, and Johnston (1969). The scale utilizes seven symbols for "negative behavior" and one symbol for "positive behavior". Table 2 shows the symbols, the classes of conduct symbolized, and the definitions of each class of conduct. We have successfully used this response code in a number of previous investigations (Baker, Stanish, and Shapson, 1971; Hislop, 1971; Baker, Stanish, and Fraser, 1972).

Each child's behavior was rated during two, separate, 5-min periods during each day of the eight-week period of the study. Each 5-min period was broken down into fifteen 20-sec time intervals. The child's conduct was observed for 20 sec and then coded with an appropriate symbol from the scale. No child could score more than one symbol per 20-sec interval, and in order to earn a relevant or positive score, a child would have to fill the requirements of that category for the full 20 sec. The number of intervals scored negatively over the total number of intervals rated (30) were expressed in percentages and constituted each child's per cent negative behavior score. The Conduct Scores Per Cent shown in Table 1 for the Token and No-Token groups were calculated on the basis of the average for the first two weeks of the program. Each child's conduct was sampled during different portions of the daily session on a predetermined random basis in order to minimize bias in the conduct sampled.

In order to minimize behavior rater bias the following procedures were undertaken. Eight raters previously unfamiliar with the response code used in the present study were employed. All raters received the same amount and kind of training in the use of the conduct rating procedures. All raters were rotated from class to class every two weeks, and each child was rated at least by four different raters during the course of the program. Further, a check on inter-rater reliability was made three times by use of videotape recordings. Midway through the program each of the eight raters independently scored the same 5-min sample of the classroom conduct of three different children using the rating procedures previously described.

Careful observation and scoring of classroom deportment of the two groups of children was made necessary by the nature of the treatment variable chosen for study. Since no explicit contingencies for reinforcement of good conduct were arranged it was quite possible that substantial improvement or worsening of the conduct of either group could occur. Under these

Table 2
General Response Code for Studying Children in a Classroom

<i>Symbols</i>	<i>Classes</i>	<i>Class Definitions</i>
X	Gross Motor Behaviors	Getting out of seat; standing up; running; hopping; skipping; jumping; walking around; rocking in chair; disruptive movement without noise; moves chair to neighbor; knees on chair.
N	Disruptive Noise	Tapping pencil or other objects; clapping; tapping feet; rattling or tearing paper; throwing book on desk; slamming desk.
Λ	Disturbing Others Directly	Grabbing objects or work; knocking neighbor's books off desk; destroying another's property; pushing with desk.
→	Aggression (Contact)	Hitting; kicking; shoving; pinching; slapping; striking with object; throwing object at another person; poking with object; biting; pulling hair.
↷	Orienting Responses	Turning head or head and body to look at another person; showing objects to another child; orienting toward another child. (Must be of 4-sec duration, not rated unless seated).
V	Verbalizations	Carrying-on conversations with other children when it is not permitted. Answers teacher without raising hand or without being called-on; making comments or calling-out remarks when no question has been asked; calling teacher's name to get her attention; crying; screaming; singing; whistling; laughing loudly; coughing or blowing loudly.
//	Other Tasks	Ignores teacher's question or command; does something different from that directed to do; includes minor motor behavior such as playing with pencil eraser when supposed to be writing; coloring while the record is on; playing with objects; eating; chewing gum.
—	Relevant Behavior	Time on task; <i>e.g.</i> , answers question; looks at teacher when she is talking; raises hand; writing assignment. (Must include whole 20-sec interval except for orienting responses shorter than 4 sec).

circumstances, day-to-day academic performance would be either enhanced or hindered and would be reflected subsequently in corresponding changes in academic achievement scores when these were obtained at the end of the program. Interpretation of the results obtained would be confounded, therefore, by the joint action of the token economy system and the effects of conduct.

2. *Academic achievement scores.* Measures of language and arithmetic achievement were obtained on three separate occasions. "Before" scores were obtained on the first day of the eight-week program and "After" scores were obtained on the last day of the program. "Before" and "After" scores were obtained in the experimental classrooms and no lessons were given on these two days. "Year Later" scores were obtained 13 months later and were obtained during testing sessions held in small offices at the children's regular schools. No tokens or other tangible

consequences were dispensed during any of the testing sessions.

The test items used to obtain measures of academic achievement consisted of nine language questions (from one test) and 35 arithmetic questions (from seven tests). The language test questions and sample questions from the seven arithmetic tests used are shown in Table 3. The type of test and its location in the instructional Teacher's Guide, which accompanies the instructional materials, are also shown in Table 3. Sample correct answers are indicated in brackets next to each question. Since correct responses to all of the test questions were wholly unambiguous and therefore easily scored, no attempt was made to measure the reliability with which examiners scored the children's answers on the tests. However, for "Year Later" scores, the precaution was taken of engaging three new examiners who had no previous contact with the children or with the

Table 3
Achievement Tests and Questions from DISTAR Instructional Materials

<i>Language Pre-Program Test: (Book A, Page 1).</i> Give this test before lesson 1. Present all of the following items to each child. For items <i>c</i> and <i>e</i> , the child is to produce the entire statement himself. (A picture of a boy in bed with eyes closed is shown.)			
<i>a. Point to the boy. What is this? (A boy.)</i>	<i>b. Is this a boy? (Yes.)</i>		
<i>c. Say the whole thing. (This is a boy.)</i>	<i>d. Is this a chair? (No.)</i>		
<i>e. Say the whole thing. (This is not a chair.)</i>	<i>f. What is the boy doing? (Sleeping.)</i>		
<i>g. Can a boy sleep? (Yes.)</i>	<i>h. Can a boy sit? (Yes.)</i>	<i>i. Is this boy sitting? (No.)</i>	
<i>Arithmetic Pre-Skills Test: (Pre-Skills Book, page 2).</i> Part <i>b</i> . Counting to a number. Listen: count to eight. What are you going to count to? Count. (The child is to count out loud, stopping at eight.) (This test has six questions.)			
<i>Arithmetic Counting to a Number Test: (Pre-Skills Book, page 130).</i>			
Pair 1. <i>a. Listen: count to four. What are you going to count to?</i>			
<i>b. Listen: count to seven. What are you going to count to?</i>			
(Child is to count to four, seven.) (This test has five questions.)			
<i>Arithmetic Counting Events in Time Test: (Book A, page 126, Test 60).</i>			
1. Every time I clap, you count.			
x . . . x . . . x . . . x. How many times did I clap? (Four).			
(This test has five questions.)			
<i>Arithmetic Object Counting Test: (Book A, page 272, Test 77).</i>			
Part 1 <i>a. Show me the objects that are in the circle. Count the objects in the circle. Do not point.</i> (The child is shown an array of seven objects printed on paper with three of them surrounded by a circle.) (This test has five questions.)			
<i>Arithmetic Concept of Equality Test: (Book B, page 165, Test 100).</i>			
Part <i>b</i> . Point to the circle with lines in the first problem. See how many we counted to on this side of the equal? Count to that many on the other side of the equal. (The child is to count to five on this question.) (This test has two questions.)			
<i>Arithmetic Counting from a Number to a Number: (Book B, page 65).</i>			
Pair 1 <i>a. Listen: You're going to count to six. What are you going to count to? Wait. And you're going to count from three. Get it going.</i> Raise your hand. After three is firm, drop your hand. (This test has five pairs of questions.)			
<i>Arithmetic Symbol Identification Test: (Book B, page 304, Test 115).</i>			
Call on each child to identify all of the symbols in one of the columns. Point to each symbol and ask: <i>What is this?</i> Wait. (seven questions).			
3	9	1°	0
5	3	8	1°
8	1°	0	8
+	5	5	5
0	8	9	3
1°	0	3	9
9	+	+	+

study. These examiners, though equally familiar with the instructional materials and procedures to those engaged for the "Before" and "After" measures, were not informed about the experimental history of the children nor the purpose of collecting the test scores. Because of practical limitations, no attempt was made to obtain daily or weekly academic achievement scores.

The Language and Arithmetic Questions Passed scores for each child were calculated simply as the total of correctly answered ques-

tions from the appropriate test(s). The Arithmetic Tests Passed score consisted of the number of tests (from the total of seven) in which each child correctly answered one or more questions.

The instructional materials were drawn exclusively from the early portions of the 2-yr DISTAR™ instructional system developed by Englemann and his associates (Englemann and Carnine, 1969; Englemann, Osborn, and Englemann, 1969). The instructional system was chosen because it provided an appropriate ve-

hicle for a token economy system, based as it is, on carefully defined behavioral objectives, explicit application of operant reinforcement, and completely structured teacher-presentation materials. The DISTAR language program consists of 180, 30-min, daily lessons in which children are taught basic language concepts. The program is divided into four segments with roughly an equal number of lessons. In each daily lesson, children are required to work on 6 to 16 tasks and to move from one segment to another, thus covering a number of concepts in each lesson contingent on a predetermined level of achievement set for each lesson. The portion of the language program used in the present study covered four areas of instruction [identity and action statements (two parts), prepositions and polars, and action statements] and introduced the children to yes-no questions, "what" questions, full statements in the form of identity statements, action statements, and "not" statements. None of our children successfully proceeded beyond the Pre-program portion of the language program in the eight-week period of the study. The DISTAR arithmetic program consists of 220 daily lessons, each with 7 to 16 tasks, and is divided into five roughly equal segments (Preskills, Books, A, B, C, and D). In the time allotted for the present study only seven areas of skill were used in arithmetic instruction including, counting to a number, counting events in time, object counting, symbol identification, equality, counting from a number to a number, and simple addition, from a total of 35 skill areas or "tracks" available).

Experimental and Instructional Procedures

Children in the Experimental (Token) Group earned poker chips (on a partial reinforcement schedule) accompanied by praise (on a continuous reinforcement schedule) from the classroom instructor contingent on giving correct answers to questions orally presented by the instructor. The questions were drawn exclusively from the DISTAR arithmetic and language teaching

materials. The questions were presented in a fashion that conformed as exactly as possible with the demands of the teaching methods, *i.e.*, every child-directed activity of the instructor was pre-programmed as much as possible. The rate of presentation and the type of questions presented were determined by the DISTAR teaching method, which demands a highly stereotyped performance from the instructor. The instructor is required to use a small set of discriminable hand and arm movements as signals for responding either from individual children or from the entire group of children in the context of a variant of the game—"Simon-says". The instructor was also required to memorize and rehearse the signals to be used before teaching, and, improvisation was not allowed. The objective was to maintain a constantly high level of attention in the children and obtain 100% participation by all of the children as well as 100% correct performance on each task (for additional details see, Englemann and Carnine, 1969, Teacher's Guide).

All correct answers were reinforced unequivocally with praise directed specifically to particular children performing correctly from moment to moment. An undetermined number of correct answers (but estimated to be about one-fifth) by the children were accompanied both by praise and a token (poker chip). Tokens were dispensed for correct responses by each child approximately every 3 to 5 min; thus allowing each child to earn from 10 to 20 tokens per day. No exact count of each child's earnings was kept. Implementation of the schedule of token dispensing relied on the practised judgments and self-pacing ability of the instructors in the present study. However, in more recent work (Rubino, Richardson, Dalton, and Hislop, *unpublished*), which guaranteed the instructors' abilities to implement a schedule (by means of radio-transmitted signals to instructors carrying "bug-in-the-ear" receivers), it was found that self-pacing of a 3- to 5-min VI schedule can be accomplished readily and reliably without support from electronic devices.

At the end of each daily session, the poker chips were exchanged for each child's preferred candies, sweets, *etc.*, at a small "store" located in a hallway adjoining the classroom. The "store" consisted of a large, gaily colored, open, wooden chest with a number of shelves upon which were displayed the items which could be bought. Since none of the children could read or count with understanding beyond "a few" or "some", each child simply received his preferred item regardless of the number of tokens in his possession. At the end of each day, the children were led out of the classroom one at a time and each was required to count and hand over each token he had in his possession to the "storekeeper" before receiving the item he had selected. All of the children were then led by their instructors to a large waiting room where they ate their rewards while waiting to be taken home in taxicabs.

Children in the Control (No-Token) Group received identical treatment except that at no time did they receive poker chips or sweets available to the children in the Token Group. Social reinforcement in the form of praise was dispensed on a continuous reinforcement schedule by the classroom instructor to each child immediately after every correct answer given to specific lesson questions. Every attempt was made to duplicate conditions in the Control Group classroom as were present in the Experimental Group classroom. This included using the same teaching materials, asking the same questions, and asking approximately the same number of questions. In addition, periodic monitoring by closed-circuit television ensured that the same rate and types of questions were asked in both groups. In the No-Token Group, as in the Token Group, incorrect answers to DISTAR questions were correlated with non-reinforcement by being ignored. Inappropriate, that is, negative behavior, was likewise ignored and never punished by the instructor. Children in the No-Token Group held their classes in the afternoon so that they were unaware of the different reinforcement system being used with the Token Group in the morning. At the end of

each day, the children in the No-Token Group likewise assembled in a large waiting room with their instructor while waiting to be taken home in taxicabs.

Classroom sessions were held three half-days per week for eight weeks. The daily program for each group consisted of two instructional half-hours separated by a half-hour recreation break. For instructional purposes, half of each group devoted the first half-hour to arithmetic lessons and the second half-hour of instruction to language lessons, with the instruction order reversed for the remaining children. Instructors and conduct raters (previously trained psychology students) were rotated four times (*i.e.*, every two weeks) between classrooms in order to minimize teacher and rater biases. No tokens were introduced during the recreation breaks for any of the children. Recreation breaks consisted of playtime with or without toys in a large waiting room or outside in a small playground equipped with the usual children's play equipment such as swings, *etc.*

At the time each child was re-tested 1 yr later, the regular teachers and principals of the children were interviewed. Information was obtained concerning possible sources of influence on scores obtained 1 yr later, *e.g.*, whether or not any of the children received additional training with DISTAR or similar teaching materials, whether or not there were systematic changes in teachers or programs involving token economy systems.

RESULTS AND DISCUSSION

Figure 1 shows the changes that occurred in the classroom conduct of the two groups. The mean per cent negative behavior is plotted against successive weeks of the eight-week program. Statistical comparisons of conduct scores (Student's "t" test, Ferguson, 1959) showed that the two groups did not differ significantly from each other at the end of the program. In addition, all but two of the children (both in the Token Group) showed a slight, and perhaps negligible, decrease in per cent negative behavior during

Fig. 1. Changes in conduct for the Token Group and the No-Token Group observed over the course of the eight-week program.

the course of the program. These findings preclude the possibility that observed differences in academic achievement scores were attributable to differences in the attentiveness or general classroom conduct of one or the other of the two groups. These results suggest further that there was no measured generalization of the effects of the token economy procedures on the conduct of the Token Group. Perhaps the absence of generalization was due to a "basement effect", *i.e.*, the children's mean per cent negative behavior scores were perhaps too close to zero from the outset of the program to show much of a decrease regardless of treatment procedures.

The reliability of the data shown in Figure 1 was established by comparisons of inter-rater judgments of conduct categories as described earlier. Comparisons were based exclusively on per cent agreement over the exact coding and categorization of types of behaviors occurring during each of the fifteen 20-sec intervals of the 5-min sample. The proportion of agreements

between each possible pair of raters (a total of 28 such comparisons) was then converted to an overall per cent agreement score. For three different reliability tests, which sampled the behavior of three different children, the inter-rater average overall per cent agreement was 91.8, 85.9, and 91.9%.

Table 4 is a summary of the results obtained before, after, and 1 yr following the eight-week program in DISTAR arithmetic and language. The individual scores for each child, identified by Group and Subject Number, are shown in the rows. Two measures of arithmetic and one measure of language achievement were obtained. The Group Means are shown below each column of scores. Comparisons for children in the Token Group of scores before with scores immediately after the program, indicated that every child except one (Subject 7) made gains in both measures of arithmetic achievement and all but one (Subject 4) made gains on the measure of language achievement. In contrast, only two of

Table 4

Academic achievement scores obtained before, after, and one year later following an eight-week instructional program in DISTAR arithmetic and language.

Group	Subject	Arithmetic Questions Passed (total 35 questions)			Arithmetic Tests Passed (total 7 tests)			Language Questions Passed (total 9 questions)		
		Before	After	Year Later	Before	After	Year Later	Before	After	Year Later
TOKEN GROUP	1	12	26	13	3	6	4	6	8	8
	2	11	29	15	3	6	5	6	7	6
	3	12	22	15	3	5	4	6	8	6
	4	0	3	7	0	2	3	6	6	6
	5	9	16	15	2	3	4	4	9	8
	6	0	6	2	0	2	3	5	7	8
	7	8	16	17	3	3	4	1	4	7
GROUP MEANS		7.4	16.9	12.0	2.0	3.9	3.9	4.9	7.0	7.0
NO TOKEN GROUP	8	0	8	1	0	3	1	4	7	7
	9	7	16	20	3	3	4	6	7	6
	10	9	8	3	3	3	2	4	8	7
	11	20	4	14	4	2	4	6	8	7
	12	0	0	9	0	0	4	5	6	6
	13	5	3	9	2	2	4	4	7	5
GROUP MEANS		6.8	6.5	9.3	2.0	2.2	3.2	4.8	7.2	6.3

the children in the No-Token Group (Subjects 8 and 9) made gains on Arithmetic Questions Passed, and only one (Subject 8) showed gains on Arithmetic Tests Passed. However, all of the children in the No-Token Group made gains in Language Questions Passed.

The results shown in Table 4 indicate that the performance of children with Down's syndrome on the DISTAR Arithmetic 1 Program can be significantly enhanced when instruction is supported by a token economy system both in terms of total questions passed ($t = 2.39$, $df = 11$, $p < 0.025$) and total tests passed ($t = 2.08$, $df = 11$, $p < 0.05$). This finding is particularly interesting because the DISTAR materials were originally designed for socially disadvantaged children in the Head Start Program (Englemann *et al.*, 1969) rather than for retarded children. The DISTAR arithmetic program, therefore, is potentially a useful new tool for teachers of retarded children.

The Token Group showed significant improvement in language scores ($t = 3.62$, $df = 6$, $p < 0.01$) during the course of the program as did the No-Token Group ($t = 4.85$, $df = 5$,

$p < 0.01$). This finding is not easily interpreted, although similar results have been reported by McDowell and his associates (McDowell, Nunn, and McCutcheon, 1969) with kindergarten children in a comparison of the effects of the look-and-say method with a cue-fading procedure. It was possible, in both studies, that conventional testing of achievement provided an insensitive measure of treatment effects. Conceivably, a test of resistance to extinction might be sufficiently sensitive to measure the extent of possible over-learning or overdetermination of responses by the experimental group, a possibility that could not be ruled out in the present study. Alternatively, perhaps children with Down's syndrome simply experience greater difficulty with arithmetic than with language, a hypothesis not inconsistent with claims to that effect made by Tredgold (1963), and thus the children require the additional incentive provided by a token economy system in order to improve.

To examine some of these possibilities, all of the children were re-tested 1 yr later on the same academic achievement measures. As during the first two test occasions, no tokens or other

tangible consequences were dispensed for correct responses to the test items. The results (Table 4) revealed that the Token Group on the whole retained its level of achievement on the academic achievement measures, although there was an unreliable decline in number of Arithmetic Questions Passed ($t = 1.87$, $df = 6$, p is n.s.). On the other hand, the No-Token Group showed a statistically significant decline in language scores ($t = 2.71$, $df = 5$, $p < 0.05$) when its scores immediately after the program were compared with scores obtained 1 yr later, while scores on arithmetic achievement showed no significant change in either direction. These results suggest that the academic gains made by the Token Group during the eight-week token economy program were somewhat more resistant to extinction than the gains in language made by the No-Token Group. It can be speculated that even relatively brief exposure to a token economy system can have relatively long-lasting effects on the achievement scores of retarded children and that the apparent initial failure to demonstrate an effect of the token economy on achievement in language does not necessarily indicate the absence of a treatment effect. A number of possible alternative explanations of the follow-up re-test results seem largely unfounded when the results of teacher and principal interviews were considered. It was ascertained that during the intervening 13 months between test and re-test, none of the children were in any token economy program and none received any additional exposure to DISTAR instructional materials other than what they received during our eight-week program. Nearly all children (except two from the Token Group and one from the No-Token Group) continued to receive some type of language training and arithmetic training. There were no systematic changes in teachers or in schools which, if present, could have selectively affected the scores of one or the other of the two groups.

REFERENCES

- Baker, J. G., Stanish, B., and Fraser, B. Comparative effects of a token economy in nursery school. *Mental Retardation*, 1972, **10**, 16-19.
- Baker, J. G., Stanish, B., and Shapson, S. A token economy for classroom control of retardates. *Ontario Psychologist*, 1971, **3**, 3.
- Bijou, S. W., Peterson, R. F., Harris, F. R., Allen, K. E., and Johnston, M. S. Methodology for experimental studies of young children in natural settings. *Psychological Record*, 1969, **19**, 177-210.
- Birnbrauer, J. S., Wolf, M. M., Kidder, J. D., and Tague, C. E. Classroom behavior of retarded pupils with token reinforcement. *Journal of Experimental Child Psychology*, 1965, **2**, 210-235.
- Chadwick, B. A. and Day, R. C. Systematic reinforcement: academic performance of under-achieving children. *Journal of Applied Behavior Analysis*, 1971, **4**, 311-319.
- Englemann, S. and Carnine, D. *DISTAR™ Arithmetic 1: an instructional system*. Chicago, Ill.: Science Research Associates, Inc., 1969.
- Englemann, S., Osborn, J., and Englemann, T. *DISTAR™ Language 1: an instructional system*. Chicago, Ill.: Science Research Associates, Inc., 1969.
- Ferguson, G. A. *Statistical analysis in psychology and education*. New York: McGraw-Hill, 1959.
- Haring, N. G., Hayden, A. H., and Nolen, P. A. Accelerating appropriate behaviors of children in a Head Start Program. *Exceptional Children*, 1969, **36**, 773-784.
- Hislop, M. W. Mod-L Project—A token economy program in a remedial setting. *Mental Retardation*, 1971, **9**, 19-20.
- McDowell, E. E., Nunn, L. K., and McCutcheon, B. A. Comparison of a programmed method of beginning reading instruction with the look-and-say method. *Psychological Record*, 1969, **19**, 319-327.
- Miller, L. K. and Schneider, R. The use of a token system in project Head Start. *Journal of Applied Behavior Analysis*, 1970, **3**, 213-220.
- Rubino, C. A., Richardson, W., Dalton, A. J., and Hislop, M. W. *Short and long-term differential effects on academic and conduct behaviors when only one set of behaviours is reinforced in token economy classroom*. Unpublished manuscript, 1972.
- Tredgold, R. F. and Soddy, K. *Tredgold's textbook of mental deficiency*. London: Bailliere, Tindall and Cox, 1963.

Received 15 July 1971.

(Revision requested 14 October 1971.)

(Revision requested 21 September 1972.)

(Final acceptance 2 November 1972.)