EFFECTS OF CONTINGENCY CONTRACTING ON STUDY RATE AND TEST PERFORMANCE¹

MARIE M. BRISTOL AND HOWARD N. SLOANE, JR.²

SOUTHEAST MENTAL HEALTH AND RETARDATION CENTER
FARGO, N.D. AND UNIVERSITY OF UTAH

A contingency contracting program designed to increase study rate and subsequent test performance was implemented with a group of undergraduate psychology students. The function of the contingency contracting program in producing increased study rate was evaluated by individual experiments with each student in an experimental contracting group. The overall effect of the program on test performance was assessed by comparing the final scores for the course earned by the experimental group with those earned by two matched control groups. A reversal procedure established that contingency contracting did significantly increase the study rate of students of a wide range of ability. However, it was selectively effective in improving the test performance of below-average students only. Study rate gains in contracted courses did not generalize to noncontracted courses. Self-recording of study time in the absence of scheduled differential consequences did not improve test performance. Study rate under no-consequence conditions varied with test schedule. For both consequence and no-consequence groups, the correlation between study time and final score for the course was only moderate.

Various studies have demonstrated the effectiveness of behavioral techniques in increasing in-class study behavior (Bushell, Wrobel, and Michaelis, 1968; Hall, Lund, and Jackson, 1968; Lovitt and Curtiss, 1969). In a college or university, however, the study behavior of concern is usually that which occurs outside the classroom, and usually requires the student to manage his own behavior.

Thus, a number of reports have appeared on the application of behavioral procedures to the development and maintenance of "study habits" (Fox, 1966; Hoopes, 1969; Ryan, 1969; Scoresby, 1969) and the general use of contingency contracting (Homme, Csanyi, Gonzales, and Rechs, 1969; Tharp and Wetzel,

1969). Clements and McKee (1968) specifically used performance contracts to increase work output of prisoners in a programmed instruction course, and demonstrated increases in work output, work efficiency, and test performance. Reinforcers included access to lunch, time in the study area, or the selection of items from a reinforcement menu. Cantrell, Cantrell, Huddleston, and Woolridge (1969) successfully applied contingency contracting in a public school setting using parents and teachers as contingency managers. In a series of experiments, Harris and Hall (*in press*) found that money incentives contingent upon improved teacher ratings, increased study time after school, or improved report card grades failed to produce significant improvements in report card grades or teacher ratings. However, when money incentives were contingent upon improved homework, quiz, and test scores, quiz and homework assignment scores improved. The specific behaviors on which reinforcers were contingent and the opportunity to emit frequent responses that could be reinforced appeared critical. Suc-

¹This study was based upon a master's thesis submitted by the senior author to the Department of Educational Psychology, University of Utah, and was partially supported by the Bureau of Educational Research and the Center to Improve Learning and Instruction of the University of Utah.

²Reprints may be obtained from Howard N. Sloane, Jr., The Bureau of Educational Research, 308 MBH, University of Utah, Salt Lake City, Utah 84112.

cessful contingency contracting using home-based reinforcement has also been reported (Bailey, Wolf, and Phillips, 1970; Cohen, Keyworth, Kleiner, and Liebert, 1971; Kirigin, Phillips, Fixsen, and Wolf, *unpublished*).

S. Moffat (*unpublished*) and J. Moffat (*unpublished*) were successful in using activity or money reinforcers to increase study time and grades of low-achieving college students in a contingency contracting framework. However, several studies have suggested that when such procedures are used, the act of recording itself may be an independent variable (McFall, 1970; Stunkard, 1970). Broden, Hall, and Mitts (1971), as well as Johnson, Martin, and White (1970), found that recording the amount of time spent on studying or other academic work increased performance with no other scheduled consequences. Sluyter and Hawkins (1972), however, suggested that with some students, concrete reinforcers, rather than feedback or social reinforcement alone, may be important in producing scholastic improvement *via* contingency contracting.

Self-monitoring (SM) has been shown to be a rather complex phenomenon. Many of its aspects have been summarized by Kazdin (*in press*). For example, studies have indicated that SM may be an inaccurate behavioral measure (Herbert and Baer, 1972) and, furthermore, may be subject to systematic biases when consequences are contingent upon SM data (Drabman, Spitalnik, and O'Leary, *in press*; Fixsen, Phillips, and Wolf, 1972; Santogrossi, O'Leary, Romanczyk, and Kaufman, 1973). Some of these studies also indicate that with training or different contingencies, accurate SM data can be obtained (Drabman *et al.*, *in press*; Fixsen *et al.*, 1971).

The present study investigated the effect of contingency contracting on study time and academic performance, as well as the effect of self monitoring alone. The relationship between study time and test performance was also assessed, as was the effect that contracting for increased study time in one course had on study rate in another course.

METHOD

Subjects and Assignments

Thirty-six undergraduate students, who volunteered from an introductory course in psychology, were ranked according to their raw scores on the first test in the course. They were blocked into groups of three, and randomly assigned to Group I, Group II, or Group III within blocks.

Experimental Design

A combined-group design and individual-subject design was used. Group I was a control group, and did not receive any of the experimental procedures. Group II was a second control group, and merely recorded and graphed study time. Group III, the experimental group, received the contracting procedure. However, these procedures were presented to each subject in Group III in a time-samples reversal design, with treatment RG being recording and graphing, and treatment C the addition of the contracting procedure. Group III also recorded (treatment R) time spent studying a second course, starting with the first exposure to the contracting procedure, a partial multiple baseline design. Test scores in Introductory Psychology were recorded for all subjects. This design is summarized in Table 1 and the procedures are detailed in later sections.

Table 1
Experimental Design

	<i>Base-line₁</i>	<i>Con-tract₁</i>	<i>Base-line₂</i>	<i>Con-tract₂</i>
Group I	RG*			
Group II	RG	RG	RG	RG
	RG	C	RG	C
Group III		R	R	R

R = recording study time alone for second course

RG = recording and graphing study time

C = contracting procedure

All groups took the test in Introducing Psychology.

*Until subjects were assigned to groups after the first test, all recorded and graphed data.

Dependent Variables

Test performance. Three multiple-choice exams were given at approximately three-week intervals during the quarter. The raw scores (number correct) for each subject and the cumulative total for all three exams were used in analysis.

Study time. Study time was the daily duration of time recorded by a subject as spent on reading, writing, typing, or project assignment for the completion of Introductory Psychology. Until the first exam, at which point subjects were assigned to groups, all subjects recorded study time. After assignment to groups, only Groups II and III recorded study time. Group III also recorded study time for a second course.

Reliability. Two checks on the reliability of study time recording for Group III were made during the second contracting period. In each, the subjects were told that the experimenter wished to assess the effect that studying in a novel location would have on study behavior. In each check, subjects were asked to study in a designated room, starting at a pre-arranged time. Subjects were given a questionnaire to fill out, after studying in this room, that purported to assess the effect of studying in a novel location (these were not analyzed). For each of these chores, subjects were paid one dollar.

In the first reliability check, subjects were asked to study in the room where they regularly met with the experimenter. This room was equipped with a one-way mirror, and an observer recorded study time from the observation room. In the second reliability check, students were asked to study in a regular classroom. An observer in the crowded hallway outside the room recorded the time the subject entered and left the study classroom.

Reliability on each check, and over both checks combined, was computed by calculating a correlation coefficient (Pearson r) between the two observers' scores. To assess the representativeness of times reported on reliability days, a correlation was similarly calculated between subject-reported times on the reliability

day and adjusted home study time on the first preceding non-zero day. Where appropriate, student-reported home study times were adjusted by 10 min, before computation, to correct for a contract criterion change.

As an informal check on complete faking of study time data, at each weekly meeting students were asked to describe the content covered in their programmed text and to show the experimenter the frames they had completed.

General Procedures

After the first meeting of the Introductory Psychology class, the volunteer subjects met with the experimenter in one group. They were told that to participate in the study they would have to record and graph study time, and were told to post their graphs where they would be visible during studying. Subjects also were told that they would have to give their graphs to the experimenter each Thursday before class, and were assured that the course instructor and his assistants would not have access to these data. Graphing procedures were demonstrated, and questions answered. Subjects were told that they would receive five dollars for recording and graphing study time daily for the next 18 days, regardless of the data reported.

Subjects were also told that at the end of approximately two weeks, they would be randomly assigned to three different groups and might receive different treatments at that time. They were told that the data reported on study time would not affect this.

Subjects were given a printed set of instructions on how to study, set up a study area, and how to use the student guide to the text they were required to purchase.

At the end of the first two weeks, after the first test, the subjects were randomly assigned to the three groups as previously described.

Specific Procedures

Group I. After the first examination, students in this group were informed that they had completed all tasks for their part in the experiment

and could collect their money for 18 days of recording and graphing. Their grades on subsequent examinations in Introductory Psychology were recorded.

Group II. This group served as a control to assess the effects of recording and graphing alone. All subjects in this group recorded and graphed study rate for the entire quarter. Graphed data were collected, first weekly and then biweekly. Praise was given at these times for submitting a completed graph. Monetary reinforcers (12 dollars) were given at the end of the quarter for submitting all graphs. No differential consequences were arranged for study rate, and subjects were occasionally reminded that the monetary consequences were not contingent upon study time. No comments on study time were made when graphs were submitted.

Group III. This group received the contingency contracting treatment, as well as recording and graphing study time for the entire quarter. This treatment was complex, and in itself comprised a single-subject research design. The sequence of conditions was the same for all subjects in the contracting group: (a) baseline₁, (b) contracting₁, (c) baseline₂, (d) contracting₂.

Baseline₁. For 18 consecutive days, subjects recorded and graphed daily study times in Introductory Psychology. This period extended three days beyond the first test. These data were submitted to the experimenter at weekly meetings.

Contract₁. In the first meeting with the experimenter after baseline, each subject signed a

study habits contract specifying that: (a) on five selected days a week he would study Introductory Psychology for the amount of time that would be specified by "task cards" given to him periodically, (b) he would honestly record and graph this study time daily, (c) he would record daily the amount of time spent studying a second course, selected as most equivalent in difficulty to Introductory Psychology, (d) he agreed to meet weekly with the experimenter, and (e) he would receive the amount of reward (money) specified by the criteria for studying Introductory Psychology.

Five signed task cards were given to each subject by the experimenter at each weekly meeting (see Figure 1). On these cards, the amount of reinforcement specified was always one dollar. For the first week, the response requirement was set as the average amount of study time reported during baseline₁. The criterion was changed each week depending upon reported study time for the previous week. If the response requirement had been met or exceeded on at least four of the five previous days, it was increased by 10 min. If it had been met or exceeded on only three of the previous five days, it remained the same for the following week. If the requirement had been met or exceeded on only one or two of the preceding five days, it was decreased by 10 min for the following week. If the study time criterion had not been met at all in the previous five days, it was reduced to half its previous value for the following week.

Immediately after completing studying for a day, the student filled in the two remaining

Fig. 1. A sample card.

TASK CARD

Name _____ Date _____ Subject _____
 Amount of Task Defined: _____ Amount of Task Completed: _____
 Amount of Reward Defined: _____ Amount of Reward Due: _____
 Date of Next Appointment: _____ Manager: _____
 Remember to bring with you to our next meeting:
 (1) your graph for Psychology 105,
 (2) your list for the second course, and
 (3) your Student Guide

blanks on a task card: "Amount of Task Completed" and "Amount of Reward Due". These cards, the graphs, and the student guide were brought to each weekly meeting.

There were never any contingencies for studying the second course.

*Baseline*₂. This condition was the same as *baseline*₁, except that subjects also continued to record their study times in the second course selected.

*Contracting*₂. This condition was identical to *contracting*₁.

For simply graphing and recording study rate for Introductory Psychology, and for merely recording study rate for the second course, subjects received 12 dollars. Subjects also received one dollar for each task card indicating that the response requirement had been met or exceeded. Social reinforcement was also scheduled for meeting the response criterion and for bringing in correctly recorded and graphed study time data. As was indicated, an additional dollar was provided as reinforcement for studying in the designated time and place for each of two reliability checks.

Because of administrative restrictions, payment could not be made until the end of the quarter. At each weekly meeting, a promissory note was filled out indicating how much was owed to the subject, and one copy was kept by the subject and one by the experimenter. At the end of the quarter, the subjects converted these to cash at the Business Office.

RESULTS

Reliability

Reliability and representativeness coefficients are presented in Table 2.

It will be noted that the coefficients for representativeness are lower than those for interobserver reliability. This is as expected, as the latter represent the agreement between data on different, rather than on the same, events. Although the correlations for representativeness are only moderate, there is no systematic bias;

Table 2

Reliability and representativeness coefficients (Pearson *r*'s) of study time data for Group III.

	<i>Reliability</i>	<i>Representativeness</i>
Check One	0.79	0.55
Check Two	0.84	0.51
Overall	0.82	0.55

of 23 pairs of data, 11 show longer study times for the on-campus day, and 12 for the off-campus day.

Subjects may have been aware that reliability was being monitored on check one. However, average reliabilities were about the same on checks one and two. Representativeness data were also approximately the same on both checks.

Group Differences in Test Performance

Final scores for each subject were obtained by summing the scores for the three tests given in Introductory Psychology. Scores for individual subjects and group mean scores are presented in Table 3. Group III, the contracting group, obtained slightly higher scores than Group I or Group II, but an analysis of variance

Table 3

Final Scores: Sum of All Three Tests

<i>Subject</i>	<i>Group I</i>	<i>Group II</i>	<i>Group III</i>
1	339	322	336
2	284	303	300
3	290	244	251
4	287	279	261
5	245	260	256
6	276	232	258
7	258	244	268
8	232	217	284
9	255	224	238
10	207	216	232
11	178	193	201
12	190	189	217
Mean	253	244	259
	$\sigma = 46$	$\sigma = 41$	$\sigma = 37$

verified that there was not a significant among-treatments effect.

A separate analysis was then performed on the better and poorer students in each group. Six students in each group were assigned to the "average or above" status, and six in each group were assigned to the "below average" status based upon their test scores on test 1, which was given while all groups were still receiving the same treatment. This generated six subgroups of students. The mean rank each of these groups obtained on each of the three tests is

- △ Group I Average or Above Students
- Group II Average or Above Students
- Group III Average or Above Students
- ▲ Group I Below Average Students
- Group II Below Average Students
- Group III Below Average Students

Fig. 2. Mean ranks of the upper half (average or above-average students) and the lower half (below-average students) in each group on the three tests. Each point represents the average of the ranks for six students.

presented in Figure 2. Of the "average or above" students, Group I, which received no treatment, had the highest ranks on tests 2 and 3, which occurred after contracting had occurred. Of these "better" students, Group III, the contracting group, showed a decline in mean test rank score over tests, while Group II, the self-monitoring group, declined on test 2 and picked up again on test 3. The results were strikingly different among the below-average students. Among these students, the contracting group (III) improved significantly with each test, while the other groups showed no improvement. This was supported in an analysis of variance (see Table 4) where the status $\times$ groups $\times$ tests interaction was significant ($p < 0.01$). The analysis was done on raw test scores.

There were 45 Monday to Friday study days during the quarter. Students in Group III did not contract during the two baseline periods; thus, the effects produced upon test performance (or the lack of results) are a function of contracting for only 60.0% (average) of the days in the quarter, and therefore may be conservative.

Study Times in Contracting Group

The mean and median study times for all 12 subjects in Group III are summarized in Figure 3. All subjects had higher mean study rates during the two contracting periods than during either of the two baseline periods (see Table 5), and t-tests between average baseline study time and average contracting study time for each subject were all significant between $p < 0.05$ and $p < 0.001$. T-tests were computed for the group between all adjacent periods, and differences were found to be significant at $p < 0.001$ for all comparisons.

The percentage of days in which no studying at all occurred decreased markedly from baseline to contracting periods (see Table 5 and Figure 4). T-tests were computed for the group between all adjacent periods, and all comparisons for number of no-study days were significant at $p < 0.001$.

Table 4
Analysis of Variance: Test Procedure

Source of Variation	Sum of Squares	DF	Mean Square	F	P
BETWEEN SUBJECTS	3108.2012	35			
Status	216.05664	1	216.05664	2.5873816	> 0.05
Groups	198.08789	2	99.043945	1.1860986	> 0.05
Status $\times$ Groups	188.93750	2	94.468750	1.1313085	> 0.05
Subject w. Status	2505.1191	30	83.503971		
WITHIN SUBJECT	5037.0625	72			
Tests	3559.1504	2	1779.5752	159.25724	< 0.01
Status $\times$ Tests	77.902344	2	38.951172	3.4858072	< 0.05
Groups $\times$ Tests	351.17773	4	87.794434	7.8568745	< 0.01
Status $\times$ Groups $\times$ Tests	378.38086	4	94.595215	8.4654881	< 0.01
Tests $\times$ Subject w. Status	670.45312	60	11.174219		

Groups = Experimental Groups I, II, or III

Status = Average or above versus below average

Tests = Tests 1, 2, or 3

Data on one representative subject (Subject 4) are presented in detail (see Figure 5). Daily study times, contract criteria, condition averages, and test days are indicated on the figure. During Baseline₁, most study time consisted of "cramming" before test 1; during baseline₂, when

no test was imminent, no studying occurred. The median study time for this subject during baseline₁ was zero, as it was for seven of the 12 subjects in this group. During contracting, study time increased and study time variability decreased for this subject, as for all subjects. By the final test, no "cramming" at all occurred,

Fig. 3. Mean and median group study rates with and without contracting. Each point represents the average (either mean or median) for 12 contracting subjects during one experimental period.

Fig. 4. The percentage of no-study days during contracting and noncontracting periods. Each point represents the average of the per cents of no-study days of 12 subjects during one experimental period.

Table 5
Mean, median, and per cent of no-study days for individual subjects (Group III) under the four experimental conditions.

Subject	BASELINE 1			CONTRACTING 1			BASELINE 2			CONTRACTING 2		
	Minutes per Day		Percentage No-Study Days	Minutes per Day		Percentage No-Study Days	Minutes per Day		Percentage No-Study Days	Minutes per Day		Percentage No-Study Days
	Mean	Median		Mean	Median		Mean	Median		Mean	Median	
1	43.8	0	61	61.8	72	20	10.0	0	89	95.6	70	11
2	40.0	0	61	66.9	62	10	10.6	0	71	123.4	75	0
3	42.0	18	44	55.4	59	10	32.9	26	44	68.1	65	0
4	29.2	0	72	80.2	60	0	0.0	0	100	60.1	76	21
5	41.7	30	44	66.2	66	0	37.3	0	57	72.0	78	7
6	41.0	30	50	54.5	60	25	34.3	30	43	114.4	90	0
7	16.7	0	60	41.7	38	10	0.0	0	100	77.7	60	33
8	19.2	0	60	46.0	38	0	33.0	30	35	72.8	60	11
9	16.9	0	83	47.8	45	15	0.0	0	100	34.5	45	44
10	49.0	30	29	55.4	48	5	23.3	30	22	86.4	78	0
11	25.1	22	44	73.3	68	0	13.3	0	67	78.3	80	0
12	31.7	0	56	49.3	40	20	14.0	0	67	78.9	50	11

Fig. 5. Daily study record for representative student (Subject 4).

which was typical of five of the 12 subjects. Subject 4 met or exceeded the study time required by his contract 86% of the time. The percentage of time that all 12 subjects met or exceeded the contract criteria ranged from 72 to 100%.

Comparison of Contracting and Non-Contracted Course

The courses chosen for study time recording in addition to Introductory Psychology included

biology, economics, general education, writing, mathematics, physiology, speech, and statistics. Data were recorded for a seven-week period for both the contracted and the noncontracted course, corresponding to the contracting₁, baseline₂, and contracting₂ periods for the contracted course. Cumulative hours of study for both courses for this entire seven-week period are presented in Table 6. Data are presented for only nine of the 12 subjects, as one withdrew from his second course and two lost data on the

Table 6

Comparison of cumulative amount of study time for contracted and non-contracted courses during a seven-week period (Group III).

Subject	Contracted Course (Psychology)	Non-Contracted Course
2	42.0	26.3
3	33.6	22.0
4	34.1	19.0
5	36.5	29.6
6	39.3	25.0
7	25.5	7.9
8	31.2	13.3
9	21.9	16.4
11	38.2	16.3
Mean	33.6 ^a	19.5 ^a

^aSignificantly different, $p < 0.001$.

second course. No subjects withdrew from or lost data for the contracted course. The mean number of hours spent on the contracted course during the seven-week period was 33.6, and on the noncontracted course, 19.5 hr. A *t*-test established that the total amount of time spent on psychology was significantly greater than the amount of time spent on the noncontracted courses ($p < 0.001$).

The average study rates in both the contracted and noncontracted courses were compared across the three experimental conditions. During the two contracted periods, the nine subjects had a higher mean study rate in the contracted course 17 of a possible 18 times, and a higher mean study rate in the noncontracted course once. During the reversal condition, when contracting was discontinued, in six of nine cases the mean study rate for the noncontracted course was higher than that in the rate in the contracted course was higher, and in one case the mean study rate was the same for both courses. Mean study times for each condition for the contracted course for these nine subjects (Figure 6) were typical of the entire group of 12 (Figure 3). The pattern in which

Fig. 6. Mean study time in contracted and non-contracted course for nine subjects.

both baseline times were lower than both contracting times was typical of all nine subjects. A similar pattern did not characterize mean study times for the noncontracted course. The declining study times shown in Figure 6 for the noncontracted course were typical of only four subjects, and in general there was a great amount of within-subject and between-subject variability for study time in the noncontracted course.

Effects of Self-Monitoring Alone

Groups II and III differed mainly in that Group III was subjected to the contracting procedure; both recorded their study times and graphed them. In Table 7, total study times

Table 7

Total times during baseline₁ and during both contracting periods for subjects from Group II (monitoring only) and Group III (monitoring plus contracting) in hours.

Pair	Group II (Monitoring)		Group III (Contracting)	
	Baseline ₁	Contracting	Baseline ₁	Contracting
1	14.8	23.1	13.1	34.9
2	5.5	14.0	12.0	40.8
3	12.5	12.9	12.6	28.7
4	23.0	55.7	8.8	34.1
5	11.0	18.0	12.5	27.8
6	10.5	22.2	12.3	35.3
7	7.3	34.9	5.0	25.5
8	20.2	21.9	5.8	18.6
9	5.4	5.8	5.0	20.7
10	6.8	11.6	14.7	31.4
11	4.7	8.2	7.5	36.2
12	8.5	21.5	9.5	28.3
Mean	10.9*	20.8**	9.9*	30.2**

* $p > 0.05$

** $p < 0.05$

during baseline₁ and for the two contracting periods are presented. A t-test indicated no difference between the groups during the baseline period, but a significant difference during contracting. Adjusted for the number of days in each period, this represents a 30% increase during contracting periods over baseline for Group II, and a 100% increase for Group III.

Figure 7 shows the average distribution of study time in psychology for all subjects in Control Group II. Each point represents the average study time for that day for all 12 subjects. Data show a pronounced scallop effect, with a marked increase in study time before each testing session and a low study rate during the intertest interval. When no differential consequences for study rate were arranged, study rate varied with test schedule.

This study pattern may be contrasted with the representative contracting graph shown in Figure 5. Although baseline segments of both graphs are similar, contracted study rates during intertest intervals are markedly higher and less

variable in Figure 5 than in comparable periods in Figure 7.

Study Time and Final Test Score

To assess the effect of study time upon test scores, a Pearson product-moment correlation coefficient was computed between cumulative study time and final test score (sum of three tests) in Introductory Psychology. Separate correlations were computed for Group II and for Group III. Each correlation was corrected for the restricted range of scores within a group by using the standard deviation of the test scores for the entire class. Each correlation was also corrected for the reliability of the study time measure. To do this, a correlation was computed between observer and subject data separately for each reliability check; these were then averaged using a z-transformation, and the resulting correlation used with both groups in the correction formula. The resulting correlations are presented in Table 8.

Table 8
Correlation Between Study Time and Test Score

Group	<i>r</i>
II	0.45
III	0.41

The correlations are of only moderate size, and approximately the same for each group.

DISCUSSION

Reliability

Questions of reliability are always raised by self-reported data. High reliability for self-reports of smoking behavior (Powell and Azrin, 1968; Thompson and Wilson, 1966) and self-medicating behavior (Azrin and Powell, 1969) have been reported. Broden, Hall, and Mitts (1971) found low reliability for day-to-day variations in studying behavior and talking-out behaviors. Fixsen, Phillips, and Wolf (1972) were unable to obtain high reliability with de-

Fig. 7. The average distribution of study time in psychology for subjects in Control Group 2. Each point represents the mean study rate for 12 subjects on one day.

linquent youths for room-cleaning behaviors without specific training in defining the responses to be scored and without specific contingencies that differentially reinforced agreement between self-reports and peer-reports. These issues are reviewed by Kazdin (*in press*).

A number of studies have shown high reliability for self-reports of study time when students studied in locations equipped with a one-way mirror (Mawhinney, Bostow, Laws, Blumenfeld, and Hopkins, 1971; S. Moffat, *unpublished*; J. Moffat, *unpublished*), or when a person in the subject's natural environment was assigned to monitor study behavior (S. Moffat, *unpublished*).

The present study used a relatively easy-to-define response, and high reliabilities between self-reports and observer reports were obtained on the two reliability days, with little difference between reliabilities obtained in the setting in which subjects might assume their study time was being recorded and that in which this was unlikely. Most subjects also showed reasonable agreement between the study times they reported on days when reliability was being checked and on the preceding day, suggesting that the data obtained on reliability days were representative of the data reported on other days and that there were no systematic inflations suggesting bias. Less than perfect reliability was not a function of the subjects reporting inflated study times; the subject reported the larger time in 13 instances, the observer in eight instances.

During meetings with the counselor, all subjects were able to answer informal questions about the text materials. In addition, half the students complained at least once that they did not believe they would be able to meet the study time criteria specified on a new task card; in six of these instances they successfully negotiated a lower time. If students were "faking" study times, it seems unlikely that they would voice these complaints.

Ideally, more reliability checks under all conditions should have been obtained. Problems of scheduling, bringing students into locations

where reliability could be checked, and the limited money available to pay for additional reliability checks prohibited this.

Contracting

It appears that contracting increased study time for all subjects. It is not clear which particular aspect(s) of the contracting procedure accounted for this increase. The monetary reinforcers, the recording and feedback, or social reinforcement from the experimenter may all have played a role. That the latter may have been significant is suggested by the fact that 91 out of 96 weekly appointments were met. However, as getting a new task card required attending the meeting, this too is confounded with other variables. Some of these variables may have interacted to increase study time, or may have interacted with the use of the student guide and the questions asked at the weekly meetings. However, from an applied point of view, contracting as an overall procedure was quite effective in modifying study time.

Generalization

Study time increases did not generalize to the noncontracted course. In addition, 11 of the 12 subjects recovered their initial baseline performance during the second baseline period, suggesting that the contracting procedure increased study time only when the experimenter maintained the experimental contingencies; other reinforcers did not gain control of the behavior. One subject did not recover baseline performance. This student reported that his grades had improved one full letter during the first contracting period, and so he set up contracts with himself during the second baseline period to maintain this gain.

Test Performance

The contracting procedure clearly improved the test scores of poorer students, but had no significant effect on the better students. This would suggest that contracting procedures be used only with "poorer" students, or that some

modifications be made for use with better students. None of the students performed so well that possible improvement was limited by a "ceiling".

"Cramming" before the final occurred with seven of the 12 students in the contracting group and for 11 of the 12 students in Group II. As three of those who did "cram" in the contracting group were "average or above" students on the initial test, and two were "below average" students, this does not appear to account for the lack of effect contracting had on the test performance of the "better" students.

Self-Monitoring

Self-monitoring alone did not improve test scores (comparison of Groups I and II), and it probably did not increase study time. The latter is not definite, as no study data were recorded when self-monitoring was not in effect; but other than test cramming days, the study times reported by this group were so low (averaging under 30 min) that it seems unlikely.

Study Time and Final Score

The only moderate correlations between study time and test performance within Groups II and III are in accordance with findings of other research. For instance, Wiggins, Pope, and Bushell (*unpublished*) reported a 0.51 correlation with college student subjects, compared with the current study's 0.45 and 0.41 correlations. As no data are available on the relationship between test measures and the course content assigned for study, or the specific skills used by students when studying, it is difficult to assess the significance of these values.

REFERENCES

- Azrin, N. H. and Powell, J. Behavioral engineering: the use of response priming to improve prescribed self-medication. *Journal of Applied Behavior Analysis*, 1969, 2, 39-42.
- Bailey, J. S., Wolf, M. M., and Phillips, E. L. Home-based reinforcement and the modification of pre-delinquents' classroom behavior. *Journal of Applied Behavior Analysis*, 1970, 3, 223-233.
- 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.
- Bushell, D., Jr., Wrobel, P. A., and Michaelis, M. Applying "group" contingencies to the classroom study behavior of preschool children. *Journal of Applied Behavior Analysis*, 1968, 1, 55-61.
- Cantrell, R. P., Cantrell, M. L., Huddleston, C. M., and Woolridge, R. L. Contingency contracting with school problems. *Journal of Applied Behavior Analysis*, 1969, 2, 215-220.
- Clements, C. B. and McKee, J. M. Programmed instruction for institutional offenders: contingency management and performance contracts. *Psychological Reports*, 1968, 22, 957-964.
- Cohen, S. I., Keyworth, J. M., Kleiner, R. I., and Liebert, J. M. The support of school behaviors by home-based reinforcement via parent-child contingency contracts. In E. A. Ramp and B. L. Hopkins (Eds.), *A new direction for education: behavior analysis*. Lawrence, Kansas, Dept. of Human Development, Univ. of Kansas. 1971. Pp. 282-308.
- Drabman, R. S., Spitalnik, R., and O'Leary, K. D. Teaching self-control to disruptive children. *Journal of Abnormal Psychology*. (in press).
- Fixsen, D. L., Phillips, E. L., and Wolf, M. M. Achievement Place: the reliability of self-reporting and peer-reporting and their effects on behavior. *Journal of Applied Behavior Analysis*, 1972, 5, 19-30.
- Fox, L. Effecting the use of efficient study habits. In R. Ulrich, T. Stachnik, and J. Mabry (Eds.), *Control of human behavior*. Glenview, Ill.: Scott, Foresman, 1966. Pp. 85-90.
- Hall, R. V., Lund, D., and Jackson, D. Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis*, 1968, 1, 1-12.
- Harris, J. W. and Hall, R. V. Effects of systematic reinforcement procedures on the performance of underachieving high school pupils. *Educational Technology*. (in press).
- Herbert, E. W. and Baer, D. M. Training parents as behavior modifiers: self-recording of contingent attention. *Journal of Applied Behavior Analysis*, 1972, 5, 139-149.
- Homme, L., Csanyi, A. P., Gonzales, M. A., and Rechs, J. R. *How to use contingency contracting in the classroom*. Champaign: Research Press, 1969.
- Hoopes, M. H. Overcoming procrastination. In J. D. Krumboltz and C. E. Thoreson (Eds.), *Behavioral counseling: cases and techniques*. New York: Holt, Rinehart, and Winston, 1969. Pp. 58-64.
- Johnson, S. M., Martin, S. B., and White, G. Self-observation as an agent of behavioral change. Eugene: U. of Oregon, 1969 (mimeo). Cited by F. H. Kanfer. Self-monitoring: methodological

- limitations and clinical applications. *Journal of Consulting and Clinical Psychology*, 1970, **35**, 148-152.
- Johnson, S. M. and White, G. Self-observation as an agent of behavioral change. *Behaviour Therapy*, 1971, **2**, 488-497.
- Kazdin, A. E. Self-monitoring and behavior change. In M. J. Mahoney and C. E. Thoreson (Eds.), *Self-control: power to the person*. Monterey, California: Brooks Cole, 1973. (*in press*)
- Kirigin, K. A., Phillips, E. L., Fixsen, D. L., and Wolf, M. M. *Modification of the homework behavior and academic performance of pre-delinquents with home-based reinforcement*. Unpublished paper presented at American Psychological Association, Honolulu, 1972.
- Lovitt, T. C. and Curtiss, K. A. Academic response rate as a function of teacher and self-imposed contingencies. *Journal of Applied Behavior Analysis*, 1969, **2**, 49-53.
- Mawhinney, V. T., Bostow, D. E., Laws, D. R., Blumenfeld, G. J., and Hopkins, B. L. A comparison of students' studying behavior produced by daily, weekly, and three-week testing schedules. *Journal of Applied Behavior Analysis*, 1971, **4**, 257-264.
- McFall, R. M. Effects of self-monitoring on normal smoking behavior. *Journal of Consulting and Clinical Psychology*, 1970, **35**, 135-142.
- Moffat, J. M. *Contingency contracting with money for students predicted to fail*. Unpublished master's thesis, University of Utah, 1972.
- Moffat, S. W. *Contingency contracting with study behavior using activity reinforcers*. Unpublished master's thesis, University of Utah, 1972.
- Powell, J. and Azrin, N. The effects of shock as a punisher for cigarette smoking. *Journal of Applied Behavior Analysis*, 1968, **1**, 63-71.
- Ryan, T. A. Formulating educational goals. In J. D. Krumboltz and C. E. Thoreson (Eds.), *Behavioral counseling: cases and techniques*. New York: Holt, Rinehart, and Winston, 1969. Pp. 70-78.
- Santogrossi, D. A., O'Leary, K. D., Romanczyk, R. G., and Kaufman, K. F. Self-evaluation by adolescents in a psychiatric hospital school token program. *Journal of Applied Behavior Analysis*, 1973, **6**, 277-287.
- Scoresby, A. L. Improving academic performance. In J. D. Krumboltz and C. E. Thoreson (Eds.), *Behavioral counseling: cases and techniques*. New York: Holt, Rinehart, and Winston, 1969. Pp. 64-69.
- Sluyter, D. J. and Hawkins, R. P. Delayed reinforcement of classroom behavior by parents. *Journal of Learning Disabilities*, 1972, **5**, 20-28.
- Stunkard, A. J. Cited by R. M. McFall. Effects of self-monitoring on normal smoking behavior. *Journal of Consulting and Clinical Psychology*, 1970, **35**, 135-142.
- Tharp, R. and Wetzel, R. *Behavior modification in the natural environment*. New York: Academic Press, 1969.
- Thompson, D. C. and Wilson, T. R. Discontinuance of cigarette smoking: "natural" and with therapy. *Journal of the American Medical Association*, 1966, **196**, 1048-1052.
- Wiggins, J. A., Pope, H., and Bushell, D. *Learning contingencies in the college classroom: a pilot study*. Unpublished final report, ERIC document number 024314, 1968.

Received 20 February 1973.

(Revision requested 19 June 1973.)

(Final acceptance 4 February 1974.)