Effects of a Multiple Baseline Design on Maintenance of Treatment Effects across Time

Bernard D. Fabry

Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Psychoanalysis and Psychotherapy Commons

Recommended Citation

https://scholarworks.wmich.edu/masters_theses/2412

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
EFFECTS OF A MULTIPLE BASELINE DESIGN ON MAINTENANCE OF TREATMENT EFFECTS ACROSS TIME

by

Bernard D. Fabry

A Thesis Submitted to the Faculty of The Graduate College in partial fulfillment of the Degree of Master of Arts

Western Michigan University Kalamazoo, Michigan August 1975
ACKNOWLEDGEMENTS

I would like to thank all those people who helped me in the completion of this thesis. The cooperation received from the staff of Schoolcraft Elementary School is most appreciated. I am especially thankful for the help of Nancy Machtel, Nadine Bawkey, David Fossum, Beverly Johnson, and Paul Timmer. I would also like to express my thanks to Dr. Howard Farris, Dr. Galen Alessi, and Dr. Brian Iwata for their advice and constructive criticism as my committee members. A special thanks goes to my wife, Pam, for her friendship and help throughout.

Bernard D. Fabry
INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
MASTERS THESIS

FABRY, Bernard Dean
EFFECTS OF A MULTIPLE BASELINE DESIGN ON MAINTENANCE OF TREATMENT EFFECTS ACROSS TIME.

Western Michigan University, M.A., 1975
Psychology, clinical

Xerox University Microfilms, Ann Arbor, Michigan 48106
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>METHOD</td>
<td>8</td>
</tr>
<tr>
<td>Subjects</td>
<td>8</td>
</tr>
<tr>
<td>Academic Measurements</td>
<td>8</td>
</tr>
<tr>
<td>Observations</td>
<td>9</td>
</tr>
<tr>
<td>Reliability</td>
<td>9</td>
</tr>
<tr>
<td>Experimental Procedures</td>
<td>10</td>
</tr>
<tr>
<td>Classroom baseline</td>
<td>10</td>
</tr>
<tr>
<td>Experimental baseline</td>
<td>11</td>
</tr>
<tr>
<td>Feedback and points for accuracy</td>
<td>12</td>
</tr>
<tr>
<td>Feedback and points for rate</td>
<td>13</td>
</tr>
<tr>
<td>Feedback and points for accuracy and rate</td>
<td>14</td>
</tr>
<tr>
<td>Bonus for accuracy and rate</td>
<td>14</td>
</tr>
<tr>
<td>RESULTS</td>
<td>16</td>
</tr>
<tr>
<td>Reliability</td>
<td>16</td>
</tr>
<tr>
<td>Accuracy and Rate</td>
<td>17</td>
</tr>
<tr>
<td>Attending and Finger Counting</td>
<td>24</td>
</tr>
<tr>
<td>Workbook Assignments</td>
<td>31</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>34</td>
</tr>
<tr>
<td>Accuracy and Rate</td>
<td>34</td>
</tr>
<tr>
<td>Attending</td>
<td>37</td>
</tr>
<tr>
<td>Finger Counting</td>
<td>38</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
BIBLIOGRAPHY
The issue of whether a behavior change, produced by some environmental manipulation, can be maintained across time has become a concern in recent years. Typically, behavior changes tend to be transient. When treatment procedures are withdrawn behaviors usually revert to baseline or near baseline levels. For example, O'Leary and Becker (1967) contacted the teachers of children who had previously been in a special token program. Disruptive behavior was found to be greater than during the token program, although somewhat lower than prior to the program. Similarly, Walker, Mattson, and Buckley (1969) reported data collected on two groups of children at three- and six-month intervals following participation in token programs. Changes in task-oriented behavior were not maintained when the children returned to their regular classrooms.

It should be expected that, in general, behavior changes would not be maintained following abrupt removal of treatment conditions. Many researchers have used the within-subject reversal design to demonstrate that treatment procedures were functionally related to behavior changes (Kazdin, 1973 and 1975). A return to or near baseline levels was actually desired. In fact, Bijou, Peterson, Harris, Allen, and Johnston (1969) recommended using short treatment periods to facilitate obtaining a reversal of effects. In light of studies where behavior changes were not maintained, it appears that maintenance should be planned. As Baer, Wolf, and Risley (1968) have emphasized, maintenance is a characteristic of applied behavior.
analysis which should be explicitly examined and programmed, rather than expected or lamented.

A variety of strategies have been used to ensure that behavior changes were maintained. Programming the environment in which maintenance was expected has been reported by some researchers. Patterson and Brodsky (1966) treated a five-year-old child with multiple problems at home and in school. To program for maintenance peers were rewarded for reciprocating positive social interactions, and the parents were trained to use contingent rewards at home. Those procedures were effective in maintaining improved behavior both at school and at home following treatment. Walker and Buckley (1968) used an individual conditioning program to improve a nine-year-old student's attending. When the student returned to his regular classroom the teacher was instructed to provide intermittent reinforcement for high rates of attending. In another study Lovaas, Koegel, Simmons, and Long (1973) trained parents to provide follow-up treatment at home for their autistic children. Evaluations were conducted two to four years after termination of the children's initial treatment program. Children whose parents were trained largely retained their gains or continued to improve.

Walker and Buckley (1972) documented long term maintenance effects for three experimental strategies following treatment in a classroom token program. The maintenance strategies included peer reprogramming, equating stimulus conditions between the treatment program and regular classrooms, and teacher training in behavior management techniques. A control group was included for comparison.
Appropriate behaviors for the peer reprogramming and equating stimulus conditions strategies were reported as significantly greater than for the control group. The teacher training strategy was not significantly different from the control group. In a later study, Walker, Hops, and Johnson (1975) compared the performance of two groups of students after treatment in a special classroom setting. Following treatment all students were returned to their regular classrooms. To program for maintenance the teachers of one group were given training in behavior management techniques and provided with weekly consultations. The other group of students served as a control and their teachers received no training or assistance. For this study appropriate behaviors were reported as significantly greater for the first group during the subsequent academic year.

Other tactics used to facilitate maintenance of behavior changes have included gradually withdrawing components of a treatment program and/or substituting reinforcers which "naturally" occur in a given setting. In one study (O'Leary, Evans, Becker, and Saudargras, 1969) an extensive token program was faded from a classroom by switching to reinforcers existing within the setting such as stars, occasional pieces of candy, and teacher approval. Follow-up data indicated sustained changes in disruptive behavior for three of five target children. In another study (Chadwick and Day, 1971) a point system with tangible backup reinforcers was initially combined with social reinforcers dispensed by the teaching staff. When behavior gains were obtained the point system was terminated while social reinforcement was continued. The results showed that high work
rates and accuracy scores were maintained following removal of the point system, but amount of time at work returned to baseline levels.

Kazdin and Polster (1973) used a token reinforcement program to improve two adult retardates' social interactions with peers. As a final phase of the program one subject was switched to intermittent token reinforcement while the other subject received continuous reinforcement. Five weeks after the program was terminated only the subject who had previously received intermittent reinforcement continued to interact socially with his peers. In another study (Jones and Kazdin, 1975) educably retarded children received token reinforcement in a special education classroom. When behavior changes had been obtained the students were switched to intermittent token reinforcement, and teacher and peer praise was added. Then, as a final phase of the program, tokens were withdrawn while teacher and peer praise was continued along with reinforcers which existed in the setting such as movies, ice cream, and extra recesses. Follow-up data, collected four months after all phases of the program had been terminated, showed maintenance of behavior changes.

Reisinger (1972) combined response cost and social reinforcement procedures to modify the behavior of an institutionalized adult patient who exhibited excessive rates of crying and no smiling. After 16 weeks the treatment procedure was reduced to social reinforcement for smiling only. Fines were eliminated. The results showed low rates of crying but frequent smiling responses during the last four weeks of the study. In another study with adults as
as subjects (Hall, Hall, Borden, and Hanson, 1975) treatment consisted of 12 weeks of training in self-management for weight loss. Three post-training sessions, spaced over another 12 weeks and including contact with a therapist, were scheduled for ten subjects. Results indicated that these subjects continued to lose weight or showed no weight gain during the post-training period.

As stated previously, many researchers have used the within-subject reversal design to demonstrate reliable control of behavior changes. That design, however, does not appear to be conducive to maintenance of such changes since a return to or near baseline levels is necessary. Another within-subject design available to researchers which may actually facilitate maintenance of changes is the multiple baseline design across behaviors. In fact, two studies employing a variation of this design have documented maintenance of behavior changes during sequential withdrawal of the treatment program. Bailey, Timbers, Phillips, and Wolf (1971) used a multiple baseline design across behaviors to assess the effects of training by peers on the articulation errors of pre-delinquent boys. Four categories of articulation errors were defined and measured over time until stable rates of responding were obtained. Training by peers was then implemented for one category of articulation errors. When stable rates of responding were again obtained the training by peers was withdrawn from the first category and switched to the second category of errors. Similarly, the switching of peer training from one category to another was repeated for the remaining two categories. Results not only showed a change in responding when training by peers
was implemented for a given category, but also showed maintenance of the change even after the training had been withdrawn and switched to another behavior category. Briscoe, Hoffman, and Bailey (1975) employed a similar research design. Nine adults, participating as policy board members in a rural community project, were trained to make behaviorally defined statements in board meetings. The training procedure was implemented separately and sequentially for three categories of problem-solving behaviors: stating a problem; finding solutions; and planning the action to the solution. In general, responding in each category increased during training for the category and was maintained when training was withdrawn and switched to another category.

Neither of the studies just discussed (Bailey et al., 1971; Briscoe et al., 1975) were designed to isolate the variable(s) responsible for the maintenance effects obtained. One possibility, however, is that the maintenance effects were partially a result of the research design employed. The purpose of the present study was to assess the effects of a modified multiple baseline design across behaviors, similar to the designs described, on maintenance of treatment effects across time.

Various researchers (Bijou et al., 1969; Kazdin, 1973) have recommended measuring behaviors that might change as a result of treatment but were not of direct therapeutic focus. For example, Ferritor, Buckholdt, Hamblin, and Smith (1972) evaluated changes in classroom attending while instituting specific contingencies for work accomplished. During the present investigation attending and
finger counting during experimental sessions, and accuracy on regular classroom mathematics assignments were assessed for possible changes as an indirect effect of the treatment procedure. Accuracy and rate of work on an experimental computation task were of direct therapeutic focus.
METHOD

Subjects

Three boys and three girls (mean age, 7.8 years) from a second-grade classroom, who, according to their teacher, were performing inadequately on mathematics assignments, served. The students' classroom was located in a small rural community (population, 1200) in southwestern Michigan.

Academic Measurements

Materials consisted of consumable workbook pages from Holt School Mathematics, Level II (Holt, Rinehart, and Winston, Inc.) and worksheets with arithmetic computation problems. Each worksheet contained 90 problems, computer generated such that there were: (1) equal numbers of addition and subtraction problems; (2) no problems requiring carrying or borrowing; and (3) 50 two-by-two-digit problem arrays and 40 two-by-one-digit problem arrays. Daily records were kept of each student's performance on the workbook pages and on the worksheets. The workbook pages were graded by the students' teacher with performance being recorded in terms of the percent of problems correct on each workbook assignment. An assignment consisted of at least two but not more than four workbook pages. The worksheets were graded by the experimenter with performance being recorded in terms of the number of problems attempted and the percent of attempted problems correct. A problem was considered attempted if there were
any numerals appearing in the answer space. The teacher required her students to attempt every problem on a workbook assignment before turning the assignment in.

Observations

Observations were conducted during the first ten minutes of each experimental session. The experimenter, equipped with a data sheet and an audible time-signalling device (Quilitch, 1972), observed the students in a consecutive, sequential order for ten seconds each. Thus, the experimenter would observe Student 1 for ten seconds, Student 2 for ten seconds, and so on through Student 6, and then return to Student 1. The observation categories were as follows:

1. Attending: student looking at his own worksheet, his own fingers, or switching his gaze between worksheet and fingers. An interval was to be scored "A" only if a student attended for an entire ten-second interval. An interval was to be scored "NA" if a student was observed not attending for any portion of a ten-second interval.

2. Finger counting: student looking at his own fingers and touching or moving the fingers of either hand sequentially. An interval was to be scored "FC" if a student counted during any portion of a ten-second interval.

Reliability

To assess the accuracy of the academic measurements a reliability grader was apportioned a number of the workbook assignments and worksheets to regrade. This grader was not provided with any information about changes in experimental conditions or given any feedback on the regrading until the study was completed. By using a clear acetate overlay the experimenter also
left no marks on the worksheets when grading them. The teacher, however, generally marked which problems had been completed correctly on the workbook pages. For the workbook assignments agreements were counted only when both the teacher and the reliability grader marked the same answers as correct. For the worksheets an agreement was counted only if both the experimenter and the reliability grader marked the same problem as attempted or marked the same attempted problem as correct. In any case differences were counted as disagreements.

To assess the accuracy of the experimenter's observation records one of four reliability observers made periodic observations. The only information given to an observer was two pages of written instructions and definitions just before each reliability observation. As with the reliability grader, no other information about the study was provided. Two of the observers had not had previous training in data collection. The records obtained from reliability observations were compared interval by interval with the experimenter's records. An agreement was counted only if the experimenter and a reliability observer recorded an occurrence of the same behavior within the same ten-second interval. A disagreement was counted when either the experimenter or a reliability observer recorded an occurrence of a behavior and the other recorded a nonoccurrence of the same behavior.

Experimental Procedures

Classroom baseline
During the afternoon of each school day the second-grade teacher held a mathematics lesson for approximately 40 minutes. This typically consisted of the teacher providing a few minutes of demonstration and instructions, and then requiring all students in the classroom to complete a workbook assignment. While the students worked the teacher circulated among them and provided assistance where needed. Any students not finishing the assignment were kept in during the afternoon recess to finish it.

This procedure had been in effect since the beginning of the school year and was continued throughout the study. However, starting seven days before the first experimental session the teacher was asked to set aside the students' mathematics assignments after grading them. The experimenter picked up the assignments the same day they were completed, recorded the students' scores, obtained reliability regradings when scheduled, and returned the assignments on the following day. This procedure was followed during this and all subsequent phases of the study.

**Experimental baseline**

At 9:30 A.M., four days per week, the experimenter took the participating students to the school cafeteria for an approximately 30-minute experimental session. When the students were seated pencils and worksheets were passed out and the time-signalling device used for observations was turned on. While the students worked the experimenter observed the students and recorded occurrences or nonoccurrences of the observation categories. After exactly 15
minutes the students were instructed to stop working and the worksheets were collected. New worksheets were then passed out and the students were instructed to continue working. After ten minutes of additional work the second worksheets were also collected and the students were escorted to the playground for their morning recess period. The students were required to work on the second set of worksheets so that experimental baseline sessions would be of the same duration as all other experimental sessions. Only the first set of worksheets were saved for later grading.

**Feedback and points for accuracy**

Starting with Session 14, the students were shown a laminated chart at the beginning of each experimental session. The chart displayed an accuracy score, in terms of number of incorrect answers, from the previous day's work for each student. The students were told that they would have 15 minutes to try to beat their own score by having fewer answers incorrect. Each student was instructed to look at the chart and state what he had to do to beat his score. Pencils and worksheets were then passed out and a 15-minute timer was set. While the students worked the experimenter recorded occurrences or nonoccurrences of the observation categories. After 15 minutes the students were instructed to stop working and the pencils were exchanged for grease pencils and clear acetate overlays which were used to protect the worksheets. As the experimenter called out the correct answers for the problems each student used his grease pencil on the acetate overlay to mark and count incorrect answers. This
scheme allowed the students to grade their own answers while keeping the worksheets unmarked. When the grading was finished each student erased his old score and entered his new score on the laminated chart. If a student had beaten his previous day's accuracy score he was instructed to color in one of the blank squares opposite his score on the chart and was given a ticket for 100 points. The points were exchangeable for such commodities as candy, toys, books, or school supplies at a value of approximately ten points per one cent. The exchange took place at the end of each experimental session but in the elementary school's special education classroom. That room had an established token economy store. Students 4 and 5 had had previous experience with the store.

**Feedback and points for rate**

Starting with Session 29, the students were shown the same chart used during the preceding phase but with their scores presented in terms of rate, number of problems attempted, rather than accuracy. At the beginning of each experimental session the students were again instructed to look at the chart and state what they had to do to beat their own scores. A student could beat his score by attempting more problems than he had attempted during the preceding experimental session. After pencils and worksheets had been passed out and the 15-minute timer set, the experimenter recorded occurrences or nonoccurrences of the observation categories. When the timer signalled the end of the work period grease pencils and acetate overlays were passed out. The students were instructed to count the
number of problems attempted by counting on the acetate overlays. The experimenter did not provide any information about the accuracy of the students' answers. When the counting was finished each student erased his old score on the laminated chart and entered the new score. If he had beaten his previous day's rate score he was instructed to color in one of the blank squares opposite his score. As in the preceding phase tickets for 100 points were awarded to students who had beaten their own scores and an exchange period was provided at the end of the experimental session.

**Feedback and points for accuracy and rate**

Starting with Session 37, the laminated chart displayed a score for each student in terms of accuracy and rate, number of problems attempted correctly. As in the preceding phases the students were given 15 minutes to beat their own score by attempting more problems correctly. At the end of the work period the grease pencils and acetate overlays were passed out and the experimenter called out the correct answers for the problems. Each student used his grease pencil to count the number of problems completed correctly by writing a number next to each correct answer on the acetate overlay. Again, if a student beat his previous day's score he was instructed to color in one of the blank squares opposite his score on the chart and was given a ticket for 100 points.

**Bonus for accuracy and rate**
At the beginning of Session 48, the students were told that they would collectively receive special bonus treats if they all worked extra hard during Sessions 48 and 49, the final sessions of the study. They were also told that the quantity of treats would be large enough to allow them to share the bonus treats with classmates who had not participated in the study. Other experimental procedures were the same as during the preceding phase with the students' scores being in terms of number of problems attempted correctly. This phase was an attempt to end the study on a positive note for the students.
RESULTS

Reliability

Reliability data were obtained for 26 of 44 workbook assignments according to the reliability computation method of Quilitch (1975). The teacher and reliability grader agreed exactly on the grading of 23 assignments (88%) and agreed within ±2 answers of each other for 3 assignments (12%). Similarly, reliability data were obtained for worksheets from 31 of 49 experimental sessions. The experimenter and reliability grader agreed exactly on the number of problems attempted for 27 sessions (87%) and agreed within ±2 problems of each other for 4 sessions (13%). Both graders agreed exactly on the number of problems attempted correctly for 20 sessions (65%), and were within ±3 for 11 sessions (35%).

Reliability observers were present for 16 of 49 experimental sessions. With respect to attending, the experimenter and observers agreed exactly for one session (6%), were within ±2 of each other for 13 sessions (81%), and ±4 for 2 sessions (13%). There were a total of 60 intervals per session. Finger counting was recorded for 15 of the 16 reliability observations. The experimenter and observers agreed exactly for 8 sessions (53%), and were within ±2 for 7 sessions (47%).

A further indication of the reliability of the observation records was provided by placing the measures of the reliability observers on the appropriate figures along with the experimenter's
measures. Cofunctional reliability occurs when the measures of two observers form similar functions (Goldiamond, 1968). As can be seen in the figures reliability measures generally covaried with the experimenter's measures, thus indicating cofunctional reliability.

Accuracy and Rate

Number of problems attempted and the percent of attempted problems correct per session are presented for Student 1 in the upper two panels of Figure 1, along with mean scores for each experimental phase. During Experimental baseline the mean rate, or number of problems attempted, was 61; the mean accuracy, or percent of attempted problems correct, was 72%. With the institution of the feedback and points contingency for accuracy, rate dropped to a mean of 6, while accuracy increased to a mean of 93%. When the contingency was switched to rate accuracy dropped to levels comparable to those of the Experimental baseline phase while rate increased to the maximum possible of 90. When the feedback and points contingency was then instituted for accuracy and rate, the mean rate dropped to 15 but showed increases across sessions. Mean accuracy increased to 90%, a level comparable to that obtained with the Feedback and points for accuracy phase. During the final phase (Bonus for accuracy and rate) rate showed a substantial increase to 71 while accuracy remained above 80%.

Similar results were obtained for Students 2 and 3, as can be seen in Figures 2 and 3, respectively.

Figures 4, 5, and 6 show number of problems attempted and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 1. Percent of attempted problems correct (upper panel), number of problems attempted (upper middle panel), number of intervals scored for attending (lower middle panel), and number of intervals scored for finger counting (lower panel), per session, for Student 1. Means for each phase are provided in boxes. Experimenter measures are depicted as enclosed dots and reliability observer measures as open dots.
FIGURE 1

EXPERIMENTAL BASELINE

ACCURACY RATE ACCURACY & RATE

PERCENT CORRECT

PROBLEMS ATTEMPTED

ATTENDING

FINGER COUNTING

SESSIONS RELIABILITY-

72 93 74 90

61 6 59 15

3.0 0.7 4.2 1.5

3.0 0.2 2.8 1.3

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 2. Percent of attempted problems correct (upper panel), number of problems attempted (middle panel), number of intervals scored for attending (lower panel), per session, for Student 2. Means for each phase are provided in boxes. Experimenter measures are depicted as enclosed dots and reliability observer measures as open dots.
FIGURE 2

EXPERIMENTAL BASELINE

PERCENT CORRECT

ACCUACY

RATE

ACCURACY & RATE

PROBLEMS ATTEMPTED

ATTENDING

SESSIONS

73 99 64 98

30 6 17 33

1.1 0.4 1.0 2.0

RELIABILITY - ○

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 3. Percent of attempted problems correct (upper panel), number of problems attempted (upper middle panel), number of intervals scored for attending (lower middle panel), and number of intervals scored for finger counting (lower panel), per session, for Student 3. Means for each phase are provided in boxes. Experimenter measures are depicted as enclosed dots and reliability observer measures as open dots.
FIGURE 3

EXPERIMENTAL BASELINE

PERCENT CORRECT

ACCURACY RATE ACCURACY & RATE

PROBLEMS ATTEMPTED

SESSIONS

ATTENDING

FINGER COUNTING

RELIABILITY

100-10 80-40 60-20 40-0

93 92 44 97

7 12 45 30

1.3 2.0 4.6 3.1

0.5 1.1 0.9 0.4

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
percent of attempted problems correct per session (upper two panels) for Students 4, 5, and 6, respectively. Results were generally similar across students except when the feedback and points contingency was instituted for rate. During this phase accuracy did not show the substantial drops found with the other students. Mean accuracy was 95% for Student 4, 91% for Student 5, and 95% for Student 6.

Attending and Finger Counting

The number of intervals scored for attending and finger counting per session, for Student 1, are presented in the lower two panels of Figure 1. As can be seen both measures showed variations generally similar to the student's rate measures. During Experimental baseline attending averaged 3.0, but then decreased to a mean of 0.7 when the contingency was in effect for accuracy. Finger counting showed a similar decrease from a mean of 3.0 during Experimental baseline, to a mean of 0.2 during the Feedback and points for accuracy phase. When the feedback and points contingency was instituted for rate, attending increased to a mean of 4.2 while finger counting increased to a mean of 2.8. When rate dropped for the Feedback and points for accuracy and rate phase, attending decreased to a mean of 1.5 and finger counting decreased to a mean of 1.3.

Similar results were found for the other students with respect to attending, as can be seen in the lower two panels of Figures 2 through 6. Finger counting is not presented for Students 2, 5, and 6, since each student's mean score for finger counting never exceeded
Figure 4. Percent of attempted problems correct (upper panel), number of problems attempted (upper middle panel), number of intervals scored for attending (lower middle panel), and number of intervals scored for finger counting (lower panel), per session, for Student 4. Means for each phase are provided in boxes. Experimenter measures are depicted as enclosed dots and reliability observer measures as open dots.
Figure 4

Experimental baseline accuracy and rate data for Sessions 1 to 49.

Percent correct:
- Baseline: 69%
- Accuracy: 91%
- Rate: 95%
- Accuracy & Rate: 93%

Problems attempted:
- Baseline: 26
- Accuracy: 18
- Rate: 28
- Accuracy & Rate: 32

Finger counting:
- Baseline: 2.7
- Accuracy: 1.0
- Rate: 3.4
- Accuracy & Rate: 3.0

Reliability: ○

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 5. Percent of attempted problems correct (upper panel), number of problems attempted (middle panel), number of intervals scored for attending (lower panel), per session, for Student 5. Means for each phase are provided in boxes. Experimenter measures are depicted as enclosed dots and reliability observer measures as open dots.
FIGURE 5

EXPERIMENTAL BASELINE

PERCENT CORRECT

ACURACY

SESSIONS

PROBLEMS ATTEMPTED

ATTENDING

RATE & ACCURACY

BONUS

100
80
60
40
20
0

10
8
6
4
2
0

35
94
91
93

39
6
16
9

1.8
0.5
1.9
0.3

RELIABILITY – 0

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 6. Percent of attempted problems correct (upper panel), number of problems attempted (middle panel), number of intervals scored for attending (lower panel), per session, for Student 6. Means for each phase are provided in boxes. Experimenter measures are depicted as enclosed dots and reliability observer measures as open dots.
FIGURE 6

EXPERIMENTAL ACCURACY BASELINE RATE ACCURACY & RATE

PERCENT CORRECT

93
95
95
98

PROBLEMS ATTEMPTED

8
5
9
27

ATTENDING

0.9
0.5
0.9
2.2

SESSIONS

RELIABILITY - ○

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
0.7 per phase. The finger counting data for Student 4 is similar to the data for Student 1. However, Student 3 showed no systematic changes across phases, the means being 0.5, 1.1, 0.9, 0.4, and 0.5.

Workbook Assignments

The mean and range of percent of problems correct per phase for each student on workbook assignments completed in the classroom are presented in Table 1. As can be seen performances showed no systematic changes across phases for any student.
Table 1
Table 1
Mean and range of percent of problems correct per phase for each student on workbook assignments completed in the classroom

<table>
<thead>
<tr>
<th>Students</th>
<th>Classroom baseline</th>
<th>Experimental baseline</th>
<th>Accuracy</th>
<th>Rate</th>
<th>Accuracy &amp; rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean range</td>
<td>mean range</td>
<td>mean range</td>
<td>mean range</td>
<td>mean range</td>
</tr>
<tr>
<td>1</td>
<td>78.7 52 to 100</td>
<td>93.2 84 to 100</td>
<td>85.9 62 to 100</td>
<td>88.7 70 to 100</td>
<td>83.6 65 to 100</td>
</tr>
<tr>
<td>2</td>
<td>91.6 84 to 100</td>
<td>92.3 79 to 100</td>
<td>85.0 74 to 93</td>
<td>69.7 27 to 96</td>
<td>84.1 55 to 100</td>
</tr>
<tr>
<td>3</td>
<td>92.1 76 to 100</td>
<td>84.7 71 to 100</td>
<td>85.5 73 to 100</td>
<td>84.7 71 to 100</td>
<td>80.6 65 to 100</td>
</tr>
<tr>
<td>4</td>
<td>87.5 68 to 100</td>
<td>71.8 71 to 100</td>
<td>88.9 52 to 100</td>
<td>93.0 90 to 96</td>
<td>84.4 70 to 95</td>
</tr>
<tr>
<td>5</td>
<td>81.6 53 to 97</td>
<td>83.2 60 to 96</td>
<td>90.9 76 to 100</td>
<td>95.8 90 to 100</td>
<td>84.8 73 to 100</td>
</tr>
<tr>
<td>6</td>
<td>84.4 33 to 100</td>
<td>84.6 56 to 100</td>
<td>82.1 70 to 100</td>
<td>92.5 85 to 100</td>
<td>77.4 65 to 95</td>
</tr>
</tbody>
</table>
DISCUSSION

Accuracy and Rate

Present results did not conclusively indicate that a behavior change, produced by a treatment procedure, could be maintained when the procedure was withdrawn and switched to another behavior. Although high accuracy scores were maintained by three of the students, substantial drops in accuracy were obtained for the other three students when the treatment procedure was switched to rate measures. The design of the investigation also did not rule out alternative explanations for either the maintenance or lack of maintenance obtained.

When considering the maintenance obtained for Students 4, 5, and 6, three alternative explanations appear tenable. Lower accuracy scores might have been obtained for these students had the Feedback and points for rate condition been in effect for a longer period of time. In fact, Student 4 showed a gradual decrease in accuracy scores across experimental sessions.

Another possibility involves a consideration of response effort when proficiency at a skill has been achieved. A student may be described as proficient with the basic math facts when the student can emit correct answers to problems almost automatically, i.e., with a short response latency. As a result of repeated drill with the basic math facts during experimental sessions and in the students' regular classroom, it is possible that Students 4, 5, and 6
had become proficient with those facts when the Feedback and points for rate conditions were put into effect. In that case the response effort required for a correct answer to a problem could have become less than the response effort required for an incorrect answer. It would then have required less effort for Students 4, 5, and 6 to continue providing correct answers to problems during the Feedback and points for rate condition and their accuracy scores would not be expected to show substantial drops. This seems to be especially possible for Student 6. Throughout all phases of the study his accuracy scores were high suggesting that he was proficient with the basic math facts prior to the study. It may have required less response effort for him to provide correct answers to problems than to provide incorrect answers and his accuracy scores would never be expected to show substantial drops.

The third alternative explanation may be found in a study by Ault, Peterson, and Bijou (1968). In that study the frequency with which an elementary school child worked on academic tasks was measured. Following a baseline period the child was given attention and praise contingent upon working. A sharp increase in that behavior was observed. When reinforcement was withdrawn for 13 sessions the behavior declined somewhat but was still twice as frequent when compared to the baseline period. Apparently the behavior was being influenced by variables not under experimental control. A similar situation may have been responsible for the maintenance of accuracy obtained for Students 4, 5, and 6 in this investigation.
When considering the lack of maintenance obtained for Students 1, 2, and 3, there are at least two possible explanations. First, it had been expected that stimuli produced by correct responses would be temporally associated with reinforcement during the Feedback and points for accuracy condition, thus acquiring their own reinforcing properties. When the treatment procedure was switched to rate it had further been expected that those stimuli would continue to be differentially associated with reinforcement and would maintain correct responses. It is possible that the differential association was not maintained during the Feedback and points for rate condition resulting in stimuli produced by correct responses losing their reinforcing value. Both correct and incorrect responses could occur during the Feedback and points for rate condition with no differential feedback being associated with correct responses.

The second explanation focuses on a possible interdependence between accuracy scores and rate scores. Other investigators (e.g., Harris and Sherman, 1973; Iwata and Bailey, 1974) have reported low accuracy scores on academic work when rate was increased as a result of some experimental manipulation. It is possible that when a student is not proficient with the basic math facts, the effort required to complete problems accurately and at a high rate is considerably greater than the effort required to simply attempt problems at a high rate. Students 1, 2, and 3 may never have become proficient with the basic math facts and any conditioned reinforcers established in this study may not have had sufficient reinforcing strength for those students to compete with any increased effort required to complete problems accurately and at a high rate.
required to complete problems accurately and at a high rate.

Had high accuracy scores been obtained for most of the students during the Feedback and points for rate condition, a research design different from that reported would have been necessary. The Feedback and points for rate condition would have been left in effect for a longer period of time. In addition, subsequent conditions would have included: (1) withdrawal of the experimental manipulation from rate as well as accuracy; (2) institution of the manipulation for rate only; and (3) institution of the manipulation for accuracy only. These conditions would have been necessary to demonstrate that accuracy and rate did not naturally or automatically covary, and that maintenance of a behavior was a function of the experimental contingency in effect for that or a temporally associated behavior.

Given the conflicting results obtained with the six students in this investigation and the inability of the research design to rule out alternative explanations it is difficult to draw any useful conclusions. However, maintenance has been reported by other researchers (Bailey et al., 1971; Briscoe et al., 1975) using a research design in which a treatment procedure was switched from one behavior to another. A follow-up study would appear to be a next step. The present study suggests that interrelated behaviors such as accuracy and rate of work on a computation task should not be chosen as dependent variables for that study. Rather, behaviors which can occur concurrently and independently would seem to be ideal choices.

Attending

In an attempt to clarify the relationship between attending
In an attempt to clarify the relationship between attending behavior and academic performance, Ferritor et al. (1972) exposed elementary school students to a series of conditions in which an experimental manipulation was alternately contingent upon either correct work or attending behavior. It was reported that during the contingency for correct work conditions attending was not systematically changed. The investigators concluded that applying contingencies directly to work performance may not alter attending behavior, and that separate contingencies are needed to produce changes in attending. In a review of the study, Klein (1975) suggested that Ferritor et al's. conclusion may have been erroneous. Other research (Kirby and Shields, 1972; Klein and Mechelli, 1973) using less complex procedures have demonstrated that work contingencies can increase attending behaviors. This study adds to the evidence against the Ferritor et al. conclusion. Attending showed variations generally similar to the students, rates of work. The more problems that were attempted, the more attending observed.

Finger Counting

Only three students showed sufficient finger counting to warrant reporting of the data. Educators (e. g., Haughton, 1971) have mentioned the importance of reaching automatic and proficient levels in skills such as number writing and computation. A decrease in rate of finger counting would suggest that a student was improving in proficiency with the basic math facts. Such a decrease might be expected when rate and accuracy of work is increased. The present
investigation failed to show such an effect. Instead, finger counting either showed no systematic changes across time or generally covaried with rate of work. These results should not be considered as conclusive, however, since the students may not have obtained combined rate and accuracy performances sufficiently high enough to show an effect.

It should be noted that each student was observed for only ten ten-second intervals during ten minutes of each experimental session according to the sequential time-sampling procedure described by Thomson, Holmberg, and Baer (1974). Other researchers (Madsen, Madsen, and Thompson, 1974; Quilitch and Risley, 1973) have reported observation strategies which permit a higher density of observations to be obtained for each subject during experimental sessions. Where possible those strategies appear to be more suitable choices than the sequential time-sampling strategy used in the present study.

A final point that deserves consideration is the inadequacy of the experimental procedure as a behavior change technique. The procedure was similar to a procedure reported by VanHouten, Morrison, Barrow, and Wenaus (1974), but with a critical difference. In the VanHouten et al. study students were required to beat their own overall best score to date. In the present investigation the students were only required to beat the score obtained during the previous day's experimental session. A clear indication of an undesirable strategy which a student could adopt can be seen for Session 38 with Student 1. For that session the student did
comparatively little work. Before the session started the student informed the experimenter that she would not be doing any work that day so that she would not need to do much work during subsequent experimental sessions. By adopting a similar strategy of not performing once every few weeks a student could adjust his required level of performance downward. That would allow him to maintain low levels of performance and yet obtain most of the rewards available. The VanHouten et al. (1974) study reported only increases in levels of performance. It thus appears that an important component of the experimental procedure is requiring students to beat their best overall score to date rather than the previous day's score. Future utilization of the procedure should include the overall best score to date.

In summary, the results of the present investigation do not conclusively indicate that a multiple baseline design across behaviors followed by the sequential withdrawal of a treatment procedure can or cannot be utilized as a maintenance strategy as well as a research design. Further investigation is needed. The results do, however, suggest that a student's attending behavior in a classroom situation positively covaries with the amount of work attempted.


Hall, S. M., Hall, R. G., Borden, B. L., and Hanson, R. W., Follow-up strategies in the behavioral treatment of overweight. Behavior Research and Therapy, 1975, 13, 167-172.


O'Leary, K. D., Becker, W. C., Evans, M. B., and Saudargras, R. A.  
A token reinforcement program in a public school: a replication 
and systematic analysis. *Journal of Applied Behavior Analysis*, 
1969, 2, 3-13.

Patterson, G. R. and Brodsky, G. A behaviour modification programme 
for a child with multiple problem behaviours. *Journal of Child 

Quilitch, H. R. A portable programmed, audible timer. *Journal of 

Quilitch, H. R. A comparison of three staff-management procedures. 

Quilitch, H. R. and Risley, T. R. The effects of play materials on 
social play. *Journal of Applied Behavior Analysis*, 1973, 5, 573-
578.

Reisinger, J. J. The treatment of "anxiety-depression" via positive 
reinforcement and response cost. *Journal of Applied Behavior 
Analysis*, 1972, 5, 125-130.

Thomson, C., Holmberg, M., and Baer, D. M. A brief report on a 
comparison of time-sampling procedures. *Journal of Applied 

Walker, H. M. and Buckley, N. K. The use of positive reinforcement in 
conditioning attending behavior. *Journal of Applied Behavior 
Analysis*, 1968, 1, 245-250.

Walker, H. M. and Buckley, N. K. Programming generalization and 
maintenance of treatment effects across time and across settings. 

Walker, H. M., Hops, H., and Johnson, S. M. Generalization and 
maintenance of classroom treatment effects. *Behavior Therapy*, 1975, 
6, 188-200.

Walker, H. M., Mattson, R. H., and Buckley, N. K. Special class 
placement as a treatment alternative for deviant behavior in 
children. University of Oregon, Eugene, Department of Special 

VanHouten, R., Morrison, E., Barrow, B., and Wenaus, J. The effects 
of daily practice and feedback on the acquisition of elementary math 