The Effects of Self-Evaluation Procedures on the Numbers and Accuracy of Alphabet Letter Writing Behavior of Preschool Children

Kevin R. Coleman

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

Part of the Developmental Psychology Commons, and the Early Childhood Education Commons

Recommended Citation
https://scholarworks.wmich.edu/masters_theses/1373
THE EFFECTS OF SELF-EVALUATION PROCEDURES ON THE NUMBERS AND ACCURACY OF ALPHABET LETTER WRITING BEHAVIOR OF PRESCHOOL CHILDREN

by

Kevin R. Coleman

A Project Report
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Specialist in Education
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
August 1985
This study assessed the effects of self-evaluation procedures on the rate and accuracy of alphabet letter writing. Three preschoolers served as subjects. During Phase I, students were given instructions, and the experimenter modeled accurate letter writing. During Phase II, instructions and modeling were again presented. Additionally, the students were given a rule statement that instructed them to use a stamper and a stamp pad to stamp their paper after first writing each of a row of four letters. The results indicated that the introduction of self-evaluation procedures was associated with immediate increases in frequency of attempts, but a decrease in accuracy. This effect, however, was not maintained across sessions. The results suggest that self-evaluation procedures alone may not be capable of consistently generating educationally significant rates of behavior.
ACKNOWLEDGMENTS

I would like to thank Galen Alessi, Howard Farris, Nancy Petty, and Ron Hutchinson for their help in completing this project.

Kevin R. Coleman
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations 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 through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

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

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Coleman, Kevin Richard

THE EFFECTS OF SELF-EVALUATION PROCEDURES ON THE NUMBERS AND ACCURACY OF ALPHABET LETTER WRITING BEHAVIOR OF PRESCHOOL CHILDREN

Western Michigan University

University Microfilms International

300 N. Zeeb Road, Ann Arbor, MI 48106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS .............................................. ii
LIST OF TABLES .................................................. v
LIST OF FIGURES .................................................. vi

CHAPTER

I. INTRODUCTION ................................................. 1
   Self-Control—Background ..................................... 2
   Self-Evaluation ................................................. 4
   Self-Recording ................................................ 8
   Self-Determination of Reinforcers .......................... 9
   Rule-Governed Behavior ...................................... 10
   Relationship Between Verbal and Nonverbal Behavior .... 11

II. METHOD ........................................................ 17
   Subjects ......................................................... 17
   Setting .......................................................... 17
   Materials ....................................................... 18
   Procedure ...................................................... 18
   Phase I ........................................................ 20
   Phase II ......................................................... 21

III. RESULTS ...................................................... 23
   Reliability ....................................................... 23
   Number of Attempts .......................................... 23
   Accuracy at Letter Writing ................................... 27
   Accuracy of Self-Evaluative Responses ..................... 27

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents—Continued

CHAPTER

IV. DISCUSSION ........................................... 28
REFERENCES ............................................. 36
BIBLIOGRAPHY ........................................... 40
LIST OF TABLES

1. Sample Worksheet ........................................... 19
2. Range, Standard Deviation, and the Mean Number of Responses Observed for Each Subject During Phases I and II .................................................. 26
LIST OF FIGURES

1. The Number of Letters Attempted by All Subjects for Each Session Conducted Across Phases I and II ........ 24
CHAPTER I

INTRODUCTION

The purpose of the present study was to determine the effects of self-evaluation procedures on the rate and accuracy of letter writing behavior of preschool children, and to determine the accuracy with which preschoolers will self-evaluate their own behavior given minimal training.

There are a few studies which have documented the effects of behavior modification techniques to develop students' letter-writing skills. Salzburg, Wheeler, DeVar, and Hopkins (1971) successfully used a combination of intermittent feedback and access to play periods to increase the accuracy with which six 5-year-old kindergartners printed letters. Ballard and Glynn (1975) demonstrated that elementary school children will write letters more accurately when self-control procedures are in effect.

The use of self-control procedures seems to offer promising opportunities for attainment of long-term goals within and outside of educational settings. In terms of educational benefits, self-control procedures have possibilities for enabling children to independently develop social and academic skills (Rosenbaum & Drabman, 1979); for teaching students to manage their own behavior (Lovitt & Curtis, 1969); and without reliance on external sources of control (O'Leary & Dubey, 1979). Self-control procedures might also help to remove some of the responsibilities for classroom discipline from the teacher and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
allow that person more time for teaching (Santogrossi, O'Leary, Romanczyk, & Kaufman, 1973) and planning (McLaughlin & Malaby, 1974). What follows will be an overview of the notion of self-control, followed by a review of several studies documenting the effects of self-control techniques on the social and academic behavior of children and adolescents.

Self-Control—Background

There are varied conceptions to what self-control involves. Most behavior exemplifying self-control occurs in the absence of readily apparent explanatory variables present in the immediate environment. Brigham (1978) has suggested that this absence has historically prompted explanations focusing on mediational variables presumed to reside within the individual. As a result, self-control has often been explained in terms of will, power, soul, drive, and other similar unobservable entities. More recently, several theorists have begun to focus attention on the role of the environment in the development of personal self-control. For example, Kanfer and Karoly (1972) suggested that all behaviors labeled self-control are learned behaviors and are, therefore, under the control of environmental and self-generated variables. Skinner (1971) proposed that self-control is no different from most other types of human behaviors in that it can be explained in terms of one's learning history and relevant environmental variables. Skinner (1953) has stated that to demonstrate self-control a controlled response must be actively manipulated by a controlling response made by the individual.
Although there seem to be many benefits produced by self-controlled behavior, teaching a person to behave in a self-controlled manner may be a difficult task. As Rachlin (1974) has pointed out, human behavior is controlled by environmental events occurring over a wide time interval. However, in all likelihood it is the events which occur soon after an instance of behavior that come to control the probability of that behavior reoccurrence. The difficulty with this arrangement is that many of life's larger payoffs are events which occur in the future and are not available after a single instance of behavior. As a result, Rachlin has called self-control a "now" versus "later" issue. A person demonstrating self-control, then, would be one who activates a controlling response in order to forego small, but tempting, immediate reinforcers for larger rewards occurring temporally distant from the behavior.

Glynn, Thomas, and Shee (1973) have identified four behavioral components of self-control. The components included self-assessment, self-recording, self-determination of reinforcers, and self-reinforcement. More recently, self-reinforcement has been reanalyzed. In the process it seems to have been redefined and grouped together with self-assessment. The following paragraphs will examine this analysis. Later, several studies will be presented that have used self-evaluation, self-recording, and self-determination of reinforcers as independent variables.
Self-Evaluation

Catania (1975) noted many contradictions in the term self-reinforcement and suggested alternative ways to explain the behavior displayed when we say that a person has reinforced himself. In terms of the contradiction, Catania noted the difficulty of assuming that a controlling response, like self-reinforcement, can operate independently of some form of external control. Like Skinner (1953), Catania (1975) contended that the occurrence and maintenance of self-control behavior is ultimately related to external control. Without some source of external control, the behavior of reinforcing other behavior would not be maintained. Rachlin (1974) elaborated on this point by calling self-reinforcement a form of secondary reinforcement since the realization or termination of the long-term reinforcers would render ineffective the controlling responses operating in self-reinforcement procedures. The final analysis, according to Catania (1975), Goldiamond (1976), and Bandura (1976), was that the critical feature of the behavior we have called self-reinforcement is the student's evaluation as to the adequacy of a response, or set of responses, in relation to a prescribed standard of performance.

Malott (1980) and Kanfer and Duerfeldt (1967) proposed that accurately monitoring one's own behavior is essential to initiating useful self-regulatory behavior. Self-evaluation may be learned through imitation and modeling, in combination with incentives. It is a difficult behavior to maintain, however. As Malott (1980) has
pointed out, parents do not normally provide their children with cues to self-evaluate; they are usually more concerned with their children's overt behavior rather than whether self-evaluation had occurred. Also, the performance criteria used as the basis for self-evaluations are often too vague to promote an accurate self-evaluation. Other reasons for failure to self-evaluate may be that the individual finds it too punishing to make accurate negative self-evaluations or too rewarding to make inaccurate positive self-evaluations.

There have been many studies already conducted which have incorporated self-evaluation procedures. Those studies have been organized into three categories.

The first category includes studies which have compared the effects of self-evaluation procedures with the effects of an externally delivered reinforcement system. As an example of such a study, Wood and Flynn (1978) compared the effects of self-evaluation to adult-dispensed tokens on the room cleaning behavior of six pre-delinquent youths in a residential setting. The results indicated that both procedures were equally effective at increasing room-cleaning behavior. The self-evaluation token system, however, was more effective in maintaining the high rates of behavior during a follow-up baseline period.

The second category includes those that have introduced self-evaluation procedures, as part of a treatment package, following some type of externally delivered reinforcement system. Drabman, Spitalnik, and O'Leary (1973) used a fading procedure to transfer the
responsibility for evaluating and reinforcing student behavior from teacher to students. Students were first reinforced by the teacher using a token system. Afterward, students recorded the points awarded by the teacher and then were awarded bonus points for matching teacher ratings. Following these steps, matching was then gradually faded in four steps. At each step, successively greater numbers of a student's points were required to match the teacher's evaluation in order for reinforcement to be delivered. Lastly, students were allowed to rate their own behavior and to deliver their own reinforcers independent of evaluations made by the teacher. The researcher found that the frequency of disruptive behavior remained at low levels while academic output remained high when students were allowed to independently self-evaluate and deliver their own reinforcers. Glynn et al. (1973) followed an externally administered system with self-control procedures that included self-evaluation. When the self-control procedures were in effect, students were to indicate whether or not their behavior was on-task when auditory signals were presented. Students noted each instance of on-task behavior by making a check on a piece of paper. The checks could later be exchanged for access to time-off plus back-up reinforcers each day. During the period of time when reinforcers were delivered by the teacher, students were on-task an average of 46% of the time. Self-control procedures, however, were associated with a further increase of 20% of on-task behavior, and additionally produced more stable rates of responding. Bolstad and Johnson (1972) compared self and external evaluation procedures in the regulation of first and
second graders' disruptive behavior. For this study, the subjects were taught to self-observe their own behavior following an externally imposed system. The subjects were also taught to self-deliver reinforcers based on their self-collected data. At the same time, subjects in an external assessment group were consequated on the basis of assessments made by the adults. The results of this study indicated a significant reduction in the occurrence of disruptive behavior for both of the assessment groups as compared to a control group. Further, the authors noted that the accuracy with which subjects in the self-observation group observed the frequency of their own behavior was high despite a lack of training. Kaufman and O'Leary (1972) used response cost and reinforcement systems to decrease disruptive behavior displayed by a group of hospitalized adolescents. After low rates had been obtained, the subjects were taught to self-evaluate their own behavior in order to test the maintenance capabilities of self-evaluation procedures. During the period of time when subjects evaluated their own behavior, low rates of disruptive behavior were maintained. However, the authors reported no significant correlation between the subject's ratings with ratings made by a teacher.

The third group of studies includes those which have assessed the effects of self-evaluation without prior exposure to an externally delivered reinforcement system. As an example of such a study, Glynn and Thomas (1974) introduced self-control procedures, with self-evaluation as one of the components, to a group of third graders without prior exposure to an externally administered reinforcement system.
system. The results of this study indicated a 41% increase in on-task behavior relative to a baseline. The authors noted that the definition for on-task behavior changed throughout the school day depending on each of the different activities. To avoid confusion and to facilitate accurate self-assessment, a cueing device was added which specified the on-task behaviors for each classroom activity. This procedure was associated with a 79% increase in on-task behavior and produced steady levels of responding. Thomas (1976) extended the above findings and has shown that self-evaluation procedures can maintain behavior for long periods of time. Santogrossi et al. (1973) obtained different results. They had disruptive students in a psychiatric hospital rate their own classroom behavior following a short baseline period. This procedure did not reduce the instances of disruptive behavior. However, when teacher-determined reinforcement was incorporated, an associated 80% decrease in disruptive behavior was observed which was maintained for only a brief period by self-evaluation plus self-determined reinforcement. Additionally, the authors reported that the correlation between student and teacher evaluations, but significantly decreased when reinforcers were based on the student's evaluations.

Self-Recording

Self-recording has been described by Glynn et al. (1973) as objectively recording the frequency with which an instance of a target behavior occurs. Several studies have examined the effects of self-recording procedures on classroom behaviors. For example,
Broden, Hall, and Mitts (1971) used self-recording to temporarily decrease the number of occasions which an eighth grade boy talked-out in class and to increase the study behavior of an eighth grade girl. In the later case, the effects observed during periods when self-recording was used were again observed during a follow-up baseline. In another study, Gottman and McFall (1972) demonstrated that participation in school could be reliably altered via self-recording procedures. Seymour and Stokes (1976) successfully used self-recording and a token system to increase the work output of three of four adolescent girls lodged in a maximum security institution for offenders. Additionally, the introduction of the same intervention strategy was associated with increases in the number of comments made by the subjects which were designed to evoke praise from the institutional staff.

**Self-Determination of Reinforcers**

Self-determination of reinforcers has been defined by Glynn et al. (1973) as determination by the subject as to the nature and amount of reinforcement to be delivered contingent upon performance of a target behavior or class of behaviors. Many of the studies conducted in this area have compared the effects of letting subjects specify the reinforcement contingencies, to the effects observed when the contingencies are defined by an external source. For example, Felixbrod and O'Leary (1973) found that behavioral productivity of second-grade students who were allowed to define the reinforcement contingency was equal to the productivity of other second grade

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
students whose reinforcement contingencies were externally imposed. Glynn (1970) found no differences between the test performance of ninth grade girls who specified their own performance standards and the test performance of other ninth grade girls whose standards were externally imposed. Lovitt and Curtiss (1969) found that academic response rate was higher for a 12-year-old boy in a special class when he was allowed to determine the contingencies for reinforcement than when his teacher arranged them for him. Additionally, the results reported by Felixbrod and O'Leary (1974) indicate that academically related behavior is nearly as resistant to extinction when prior contingencies are self-specified as when the contingencies are specified by an external source.

**Rule-Governed Behavior**

The exercise of self-control seems to require a degree of rule following. In a discussion of the operant analysis of problem solving, Skinner (1969) defined rule-governed behavior and contrasted it with contingency-shaped behavior. He suggested that rule-governed behavior is behavior under the stimulus control of any statement of the relationship between a response and its contingencies. Laws, maxims, and models are examples of rules. The probability that one will follow a rule is dependent upon the degree to which the rule actually predicts the consequences. Contingency shaped behavior is behavior under direct control of the contingencies themselves. Rule governed behavior seems to be controlled by a stimulus that has no inherent motivational properties, while the stimuli operating on
contingency-shaped behavior actually do shape behavior and help determine the probability of that behavior reoccurrence. Because the nature of the stimulus is different for the two types of behavior, Skinner (1969) stated that rule-governed behavior and contingency-shaped behavior are two different operants. He noted that the rules operating as stimuli for rule-governed behavior are often general and do not produce the exact same behavior that would evolve if naturally occurring consequences were available. Rule-governed behavior is less sensitive to the consequences than is contingency-shaped behavior.

Skinner (1969) suggested that rules may be very effective when current contingencies are defective, when the contingencies are long-deferred, or when the current contingencies act to shape unwanted behavior. In short, it appears that rules can serve an important role in situations that call for self-control.

Relationship Between Verbal and Nonverbal Behavior

Rule-governed behavior has been identified as important for self-control. Rules are a form of verbal behavior. In the following, the relationship between verbal and nonverbal behavior will be examined.

Most of the early research designed to document the relationship between verbal and nonverbal behavior was done so with the assumption that a functional relationship exists between the two (Karlan & Rusch, 1982). Therefore, it was felt that nonverbal behavior could be modified by simply altering related verbal behavior. Lovaas
(1961) early on hypothesized that the presumed relationship between verbal and nonverbal behavior was due to reinforcing stimuli common to the corresponding verbal and nonverbal behavior, or because one acted as a discriminative stimulus for the other. Blackwood (1970) suggested that verbal mediation gains functional control over nonverbal behavior through a process of stimulus generalization. He hypothesized that many of the verbalizations we make to ourselves have many of the stimulus properties that are possessed by the statements made by others. Therefore, self-verbalizations assume the stimulus and reinforcing properties associated with statements or commands made by others.

Early studies designed to control nonverbal behavior focused on attempts to alter related verbal behavior. This has been called verbal conditioning by Israel (1978) and is exemplified by the work of Sherman (1964), Lovaas (1961), and O'Leary (1968). Sherman (1964) and Lovaas (1961) found that reinforcing verbal statements had a small effect on corresponding play behavior. Based on his results, O'Leary (1968) tentatively suggested that first graders who were trained to accurately use self-instructions produced fewer instances of "immoral behavior" than those students who were not trained to use self-instructions. Catania (1982) found that verbal behavior will have a greater effect on nonverbal behavior when the verbal behavior is shaped rather than instructed. Taken together, the studies by Sherman (1964), Lovaas (1961), and O'Leary (1968) indicate that increasing verbal behavior alone has only marginal effects on corresponding nonverbal behavior. It appears that most likely, verbal
behavior is under a total different set of contingencies than non-verbal behavior. The consequences for saying that one will engage in a subsequent behavior are usually pleasant. In the immediate situation, it is often very reinforcing to agree to engage in some form of behavior when asked. However, the consequences for actually engaging in the corresponding nonverbal behavior, for "doing," may vary. Corresponding nonverbal behavior may actually have associated reinforcing consequences, for example, engaging in a repetitious, laborious task, and also may have other, hidden, consequences which may involve an opportunity cost for not engaging in some other task. For example, having to engage in a repetitive, laborious task may involve a certain cost when instead one could be playing tennis or going to the beach.

More recently, a number of studies have been conducted which have attempted to bring verbal and corresponding nonverbal behavior under control of the same consequence, and to treat both as the sequence of behavior. This approach has been called correspondence training and is exemplified by the work of Bem (1967), Burron and Bucher (1978), Risley and Hart (1968), and Rogers-Warren and Baer (1976). The results obtained by Bem (1967) indicate that developing a regulatory capability of verbal behavior over nonverbal behavior is not simply a developmental problem as was suggested by Luria (1961), but rather may be a function of training. Studies conducted by Burron and Bucher (1978), Risley and Hart (1968), and Rogers-Warren and Baer (1976) indicate that verbal behavior may control nonverbal behavior if the sequence is trained directly. Further, it is
possible, after extended correspondence training, to control non-verbal behavior by only altering verbal behavior (Risley & Hart, 1968). It appears that the most productive way of training correspondence is through a say-do, rather than a do-say paradigm (Israel & O'Leary, 1973).

The present study most closely resembles the study conducted by Santograssi et al. (1973). That study introduced self-evaluation procedures following a period when disruptive behaviors were either ignored or reprimanded, and behavior that exemplified desirable classroom conduct was verbally praised. These baseline procedures may be considered typical classroom procedures in effect during the course of the day. The present study used instructions and modeling as its baseline measure. Like the Santograssi et al. (1973) study, these procedures reflect routine classroom procedures in effect during baseline.

Primarily, the present study focused on self-evaluation as an intervention strategy. From a review of the available literature, the effects of self-evaluation on behavior appear to be in question. Some of the reported studies showed that desirable changes in behavior were associated with the introduction of self-control procedures (Bolstad & Johnson, 1972; Glynn et al., 1974). However, the effects of self-evaluation alone could not be determined either because self-evaluation was not the sole intervention strategy (Bolstad & Johnson, 1972; Drabman et al., 1973; Glynn et al., 1974), or because prior association with external programs prevented a clear analysis of the behavior change capabilities of self-evaluation procedures (Bolstad &
Johnson, 1972; Glynn et al., 1974; Kaufman & O'Leary, 1972). Although the results reported by Wood and Flynn (1978) appear to be promising, the changes in behavior associated with their self-evaluation procedure were small. The Glynn and Thomas (1974) study also suggested that desirable changes in behavior may be obtained using self-evaluation procedures. Interpretation of the changes reported in that study, however, is limited due to the number of components in their self-control procedures and to the short length of time during which data were collected for the self-control phase. The results obtained by Thomas (1976) are more encouraging. The author replicated the Glynn and Thomas (1974) study and showed desirable changes in behavior while collecting data for a longer period of time when subjects were instructed to use self-control procedures. However, here again the self-control package included other components than self-evaluation, making it difficult to assess the effects of self-evaluation alone. The results of the Santograssi et al. (1973) study do not indicate that self-evaluation procedures introduced without prior exposure to external programs are effective means for changing behavior. However, the authors of that study noted that the subjects were students enrolled at a psychiatric hospital, and they cautioned their readers about generalizing from their study with hospitalized subjects to normal populations.

In general then, the present study expands on the available literature by increasing the amount of data gathered on which conclusions may be drawn concerning the effectiveness of self-evaluation procedures. Additionally, the study could provide information
regarding the accuracy with which preschoolers assess their own behavior.

In the present study, the researcher modeled accurate letter writing behavior and then instructed the subjects to do the same and to keep writing until they were told to stop. During the second phase, the Phase I conditions were continued. However, the experimenter added instructions for subjects to put an ink stamp on their paper after they had written a row of four alphabet letters. It was anticipated that allowing students to overtize self-evaluations of their work would produce increases in the rate with which the subjects wrote the forms of alphabet letters while accuracy would decrease. It was also expected that the subjects would self-evaluate their work with a high degree of accuracy.
CHAPTER II

METHOD

Subjects

The three subjects were preschoolers enrolled at the Child Development Center in Kalamazoo, Michigan. The subjects were all 4-year-old males. Each had been receiving instruction using the Direct System for Teaching and Remediation (Distar) curriculum. As part of that program, the subjects were required on a daily basis to practice writing the forms of letters. At the time the study was initiated, the three subjects had completed from 35 to 60 lessons from the Distar Reading I curriculum. Subjects were chosen on the basis of a report by their classroom leader. That person reported that each of the subjects wrote the forms of letters relatively accurately, but at a low or variable rate.

Setting

The study was undertaken at the Child Development Center in Kalamazoo, Michigan. The Child Development Center was an affiliate of Western Michigan University which offered preschool services to the local community. Sessions were conducted in a vacant classroom. The dimensions of the classroom were approximately 20 feet by 30 feet. In the room were located some shelves with toys and books, a large desk for the classroom leader, a play kitchen area, and a table.
measuring 5 feet in diameter with four small chairs. The classroom was similar to the three classrooms that were used by the Center to deliver their services.

Materials

The materials used during the course of the study included a stopwatch, beginner writing pencils, sheets of paper with the forms of the alphabet letter "m" dittoed on them (see Table 1), three stamp pads, and three stampers. The students used this material to write the letter "m" and to stamp their papers. From the subjects' finished worksheets it was possible to tally the number of letters attempted and the percentage of accurately written letters following each session. During the course of each session a separate data sheet was used to record the accuracy with which the students self-evaluated.

Procedure

Two phases were used to complete this study. Ten-minute sessions were conducted twice a day, 5 days a week, for a total of 50 sessions. From the subjects' worksheets collected from each of the sessions, two measures were taken. The two measures were number of attempts per 10-minute sessions and percentage of accurately written letters. An attempt was recorded whenever a subject wrote inside the outlining blue area of an alphabet letter (see Table 1), and the form they drew contained two humps similar to that contained by the lower case "m." To record the percentage of letter forms written
<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Worksheet</td>
</tr>
</tbody>
</table>

[This sample worksheet has been reduced to 74% of its original size.]
accurately, a binary point system was used. If the subjects stayed on or within the blue outlining area (see Table 1), they received one point. Each of the downstrokes was allowed to exceed the page lining by 1/8 of an inch. If the subject's attempt did not meet the above criteria, a score of zero was recorded for that attempt. Letter-writing accuracy for each session was calculated as follows: \( \text{number of letters accurately written/number of letters attempted} \times 100. \)

Reliability checks were made by an independent observer every fifth session for a total of 20\% of the sessions conducted. Reliability was calculated as follows: \( \text{agreements/agreements plus disagreements} \times 100. \)

**Phase I**

During Phase I the subjects and the experimenter sat at small chairs situated around a small white table located in an unoccupied preschool classroom. The subjects were randomly seated at the table for the first session. For each of the following sessions the subjects were assigned the same seats. During Phase I the experimenter used instruction and modeling to demonstrate appropriate responding. A session was initiated when the experimenter said, "Today we are going to do some writing. Watch me, I am writing on the lines and in the blue area." Following this prompt, the experimenter modeled accurate letter writing by completing a row of four letter forms. While writing each of the letters, the experimenter repeated the rule statement for staying on the lines and in the blue area. Non-examples of accurate letter writing were not presented. After the
experimenter completed a row of letters, he said, "Now, I want you to do the same thing, and I want you to keep on writing until I tell you to stop." After worksheets and pencils were passed out, the subjects were told to begin and the stopwatch was started. During this time the experimenter remained at the table with the subjects but did not interact with them other than to provide extra worksheets when necessary.

Phase II

During this phase the experimenter again sat at the table without interacting with the subjects other than to provide initial modeling, instructions, and additional worksheets when necessary. Sessions were initiated when the experimenter said, "Today we are going to do some writing. Watch me, I am staying on the lines and in the blue area." As he did during Phase I, the experimenter modeled accurate letter writing by completing each of a row of four letters while repeating the rule statement for staying on the lines and in the blue area. Additionally, the use of an ink stamper was used. When all of the letters of the first row had been completed, the experimenter said, "See, I have completed a whole row of letters and now I get a stamp on my paper. Remember, we only stamp our paper after completing a whole row of letters." Using a stamper and a stamp pad, the experimenter then modeled stamping the worksheet. Then the experimenter continued by writing the first letter in the second row. Afterwards he asked, "Do I stamp my paper now? No." The same script was used for the following letters until the last
letter of the row had been completed. At that time the experimenter asked, "Do I stamp my paper now? Yes." The experimenter followed by again modeling accurate self-evaluation by stamping his paper. Work-sheets, stamp pads, and pencils were passed out. Subjects were allowed to use one of the three stampers which they selected from a display of the designs each made. The subjects were then instructed to begin writing, and to keep writing until they were told to stop.

Several criteria were used to judge the accuracy of the self-evaluations. Most of the criteria focused on eliminating the possibility that the subject incorrectly self-evaluated. Those possibilities were as follows: If a subject stamped his paper before completing a row of letters, if he failed to stamp his paper after completing a row of letters, or if he gave himself more than two stamps after completing a row of letters, an instance of inaccurate self-evaluation was recorded for that row. An accurate self-evaluation was recorded whenever the student met none of the inaccuracy criteria and stamped his paper not more than twice immediately after attempting all four letters of a row. The accuracy with which a subject self-evaluated was calculated as the frequency of accurate self-evaluation/opportunities to self-evaluate x 100. Reliability for this measure was taken by an independent observer every fifth session for a total of 20% of the sessions conducted. Reliability was calculated using the following formula: agreements/agreements plus disagreements x 100.
CHAPTER III

RESULTS

The results indicate that the number of letters attempted and the accuracy of the letters that were written deteriorated rapidly for all three subjects in the absence of externally imposed consequences. The data further suggest that preschool boys will not consistently self-evaluate their behavior accurately when no specific consequences for this behavior are externally imposed.

Reliability

The reliability for number of attempts and accuracy of letters written was 100% and 98%, respectively. The reliability for the accuracy with which the subjects self-evaluated their own behavior equaled 94%.

Number of Attempts

Figure 1 shows the number of letters attempted by all subjects during each of the 10-minute sessions. An attempt was scored whenever a subject wrote inside the outlining blue area of an alphabet letter, and when the form drawn contained two humps similar to those contained in the lower case letter "m." From Figure 1, it can be seen that the pattern of responding was very similar for all three subjects across both experimental conditions. During Phase I, the number of attempts made by each subject during the initial session
Figure 1. The number of letters attempted by all subjects for each session conducted across Phases I and II.
was relatively high and variable as compared to the later sessions when few or no attempts were made by any of the subjects.

Phase II was initiated after 22, 32, and 43 sessions for Subjects 1, 2, and 3, respectively. The trend of the data collected during Phase II is very similar to that for Phase I for Subjects 1 and 2. Following the introduction of self-evaluation procedures during Phase II, the number of attempts for Subjects 1 and 2 was initially high and variable. These high rates, however, were not maintained; later sessions were associated with rapid declines in the number of letters that they attempted. The introduction of self-evaluation procedures for Subject 3 was associated with a very slight and temporary increase in the number of letters that he attempted.

Table 2 shows mean number of attempts, the standard deviation, and the range with which each subject responded during Phases I and II. The data presented in Table 2 confirm that responding for all subjects was highly variable during Phase I as indicated by the wide ranges and high standard deviations. Also, responding for each of the subjects during this phase was more frequent during the initial sessions as indicated by a comparison of first and second half responding for all subjects.

During Phase II, responding for all subjects remained variable, but was not as variable as the responding observed during Phase I. For each case, the standard deviations are lower, and the range of attempts was shorter for each subject during Phase II as compared to Phase I. The data presented in Table 2 indicate that the mean number of attempts decreased for Subject 1, increased for Subject 2, and
Table 2

Range, Standard Deviation, and the Mean Number of Responses Observed for Each Subject During Phases I and II

<table>
<thead>
<tr>
<th></th>
<th>Phase I</th>
<th>Phase II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0-90</td>
<td>0-60</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>Mean number of attempts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>Sessions 1-11 (first half)</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Sessions 12-22 (second half)</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Sessions 23-36 (first half)</td>
<td></td>
<td>37</td>
</tr>
<tr>
<td>Sessions 37-50 (second half)</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td><strong>Subject 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0-84</td>
<td>0-48</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>Mean number of attempts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Sessions 1-16 (first half)</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Sessions 17-32 (second half)</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Sessions 33-41 (first half)</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>Sessions 42-50 (second half)</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td><strong>Subject 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0-58</td>
<td>0-14</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Mean number of attempts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Sessions 1-22 (first half)</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Sessions 23-43 (second half)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sessions 44-47 (first half)</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Sessions 48-50 (second half)</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
remained the same for Subject 3 after changing from Phase I to Phase II conditions. However, as is shown in Figure 1, a comparison of first with second half responding for Subjects 1 and 2 indicates a rapidly decreasing trend across sessions. Figure 1 also suggests that Subject 3 responded at a very low rate throughout the study.

Accuracy at Letter Writing

For each session, the accuracy with which the subjects wrote the alphabet letter "m" was recorded. The general trend of the data indicates that the subjects more accurately wrote the letter "m" during Phase I than Phase II. As compared to Phase I, accuracy decreased during Phase II by 32%, 25%, and 21% for Subjects 1, 2, and 3, respectively.

Accuracy of Self-Evaluative Responses

During Phase II, the accuracy with which the subjects followed the given rule for self-evaluating their own behavior was recorded for each session. The mean accuracy of self-evaluative response was 75%, 75%, and 40% for Subjects 1, 2, and 3, respectively.
CHAPTER IV
DISCUSSION

The results indicate that the number of letters attempted and the accuracy of the letters that were written deteriorated rapidly for all three subjects in the absence of externally imposed consequences. The data further suggest that preschool boys will not consistently self-evaluate their own behavior accurately when no specific consequences for this behavior are externally imposed. Thus, it appears that similar strategies may be effective in the long-run only if external consequences are provided for making accurate self-evaluative responses and if the target behavior (writing) itself is occasionally reinforced by some external source.

The results of this study directly support other research and early theoretical discussions related to self-control. Skinner (1953) stated that behavior in general, including behavior one would classify as exemplifying self-control, is ultimately related to external environmental control. In the present study, no external feedback was given during Phase I or Phase II for the number and accuracy of letter-writing attempts, nor was any feedback given during Phase II based on the number of stamps each child accurately delivered to his paper. Within each phase, the number of letters attempted decreased rapidly for each subject, and over the course of the study, so did the accuracy with which letters were written. Additionally, there was a large amount of inter-subject variability.
in accuracy of self-evaluative responses. These findings seem to support much of the theoretical discussion of self-control, including Skinner's (1953) notion of the importance of environmental control, and would suggest that future research investigating self-control strategies include provisions for generating and maintaining the behaviors referred to by Skinner as controlling and controlled responses.

The results of this study also support many of the research studies already conducted. Santogrossi et al. (1973) found that self-evaluation procedures alone were not effective in changing or maintaining desirable rates of behavior. Several other studies (Bolstad & Johnson, 1972; Glynn et al., 1973; Glynn & Thomas, 1974; Kaufman & O'Leary, 1972; Thomas, 1976; Wood & Flynn, 1978) have documented the effectiveness of using self-evaluation procedures in conjunction with other intervention procedures, such as a token economy. The results of the present study, together with those studies just cited, underscore the need for additional interventions designed to generate or maintain target responses when self-evaluation procedures are used.

The present results are also consistent with the results obtained related to the accuracy with which subjects self-evaluate their own behavior. Bolstad and Johnson (1972) found that 75% of the ratings made by the subjects in their study was within a range considered to acceptably resemble the evaluations made simultaneously by an outside agent. In the present study, two subjects self-evaluated their own behaviors with 75% accuracy, and a third with 40% accuracy.
These results from this study, however, should only tentatively be interpreted, as there are a number of qualifiers to the data.

First, the measures taken on accuracy of self-evaluation were incomplete. A more completely descriptive way to measure accuracy of self-evaluation would have included for each subject a frequency count of the number of stamps delivered appropriately in relation to the number of stamps delivered during each session. It was hoped that this type of measure could be taken and reported in the results of the present study. However, during the later sessions of the experiment, the subjects spent a large portion of time simply stamping their papers. Under these conditions, the stamps did not remain discrete, and it became impossible to accurately count them. Therefore, the data for accurate self-evaluation is not complete. Future research using similar procedures should attempt to insure that each instance of overt self-evaluation can be accurately counted. The second qualification to the results of the present research is related to the length of time that subjects were allowed to self-evaluate. The length of time that the subjects were exposed to self-evaluation conditions was relatively short and, therefore, did not allow for an assessment of accuracy of self-evaluations over a large number of sessions.

In general, the small number of sessions conducted during Phase II represents the largest limitation of this study. Although the trend of the data for each subject during Phase II suggested initial increases in frequency of attempts followed by slowly declining rates, the total mean frequency of attempts remained the same for one
subject and actually increased for another from Phase I to Phase II. Had the number of sessions conducted for each subject been increased for Phase II, it would have been possible to assess the effects of self-evaluation on frequency of attempts over a longer period of time. However, it should be noted from Table 2 that the total mean frequency of attempts during Phase II was 23, 19, and 4 for Subjects 1, 2, and 3, respectively. Independent of the declining rates suggested by the trends in Figure 1 and the data in Table 2, this researcher does not believe that the total mean frequency of attempts collected during Phase II represents behavior that one would consider to be educationally significant, suggesting that the opportunity to self-evaluate had a very weak effect on the desired behavior.

A clarification of the results presented in Table 2 is necessary. The data presented for the range, standard deviation, and mean frequencies of attempts were presented to numerically support the visual effects of Figure 1. The data were not presented to assist in providing an inter-subject comparison of performance. Such a comparison would be inappropriate given the experimental design used for this study. When using a design like the one used in this study, it is not possible to make valid comparisons of performance since all subjects are not always exposed to the same experimental conditions at the same time. In fact, some of the inter-subject differences in responding indicated by the data in Table 2 may primarily be accounted for in terms of the requirements of the design. For example, the data in Table 2 indicate a fairly large difference between the total mean frequency of attempts during Phase I for
Subjects 1 and 2. Inspection of Figure I, however, would probably lead one to conclude that Subjects 1 and 2 demonstrated similar rates of behavior and similar trends until Session 23, when a phase change was made for Subject 1. The differences documented in Table 2, related to the total mean frequency of attempts for Subjects 1 and 2, can be accounted for in terms of the deflating effect that extending Phase I conditions had on the total mean frequency of attempts for Subject 2. Furthermore, an inter-phase comparison of the standard deviations would, in this case, be misleading since the subjects did not respond as frequently during any Phase II session as they had during many of the sessions during Phase I. This lower level of responding imposed restrictions on the size of the standard deviations that could be obtained.

Some discussion should occur as to the function of the stamps used in the present study, as well as some speculation as to why the subjects stamped their paper relatively accurately during initial, but not during the later sessions of Phase II. All subjects stamped their papers during each of the sessions conducted during Phase II. One could be fairly safe to say then, that putting a stamp on their paper was a reinforcing activity for these subjects. The stamps probably acquired their reinforcing function before the study was initiated. This may have occurred through prior association with having a stamp or a small gummed star put on their papers, paired with praise delivered by their classroom leader or other classroom aides at the Center. The obvious difference in the subjects' use of the stampers across sessions in Phase II was the way that subjects
followed the given rule for evaluating their own behavior. During initial sessions the subjects followed the given rule for accurately assessing their behavior; during the later sessions, they did not. Instead of completing all four letters in a row and then stamping their papers as they did relatively well during the initial sessions of Phase I, the subjects stamped their papers without attempting many letters. One can only speculate as to why this occurred. This researcher would suggest that this change occurred as a function of the changes in the contingencies of reinforcement. One of the stated objectives of the Center, as part of social skills training, was to increase the skill of the children to follow rules and directions. As part of that training, the children at the Center received intermittent feedback from their classroom leader and aides. One would expect then, that the subjects would initially demonstrate skill at following rules and directions for a period of time in the research setting despite a lack of immediate feedback for doing so. As feedback was continually withheld, however, the subjects' learning history became inoperative, and the consequences of the immediate environment came to control stamping behavior. The consequences for repeatedly stamping the worksheet included the production of a reinforcing product on the environment, and attention delivered by other subjects. This analysis would support Skinner's (1953) notion that all behavior related to self-control is ultimately related to environmental variables. In terms of the present research, stamping a worksheet after completing a row of four letters would be called a controlling response as defined by Skinner. In the absence of
environmental variables designed to generate the controlling and controlled responses, the whole self-control process becomes inoperative.

To continue the analysis of the function of the stamps, it is also believed by this researcher that the behavior of stamping also, initially, had a discriminative function. When the behavior of the subjects appeared to be under control of the rule for accurate self-evaluation, the subjects would quickly write each of the four letters in a row, stamp their paper, and then would continue to write. Apparently, at that time the contingencies stated in the rule statement were controlling the stamping behavior of the subjects. As part of that rule, and given that stamps had a reinforcing value, an instance of reinforcement was not available after the paper had been stamped once. Under these conditions, a stamped paper might have served as a discriminative stimulus for an occasion upon which writing a row of letters would be followed by a reinforcing event. This discriminative function probably was weak. After repeatedly stamping the worksheet without first writing, and without any consequence from the researcher, the contingencies associated with stamping the worksheet (i.e., impacting the environment with a reinforcing permanent product and feedback from other subjects) came to control stamping behavior, and observation of a newly printed stamp became a discriminative stimulus for engaging in the behavior required to put another stamp on the paper. Within this framework, it is easy to understand why the relatively fast rates observed when subject's stamping behavior coincided with that suggested by the rule for
accurate self-evaluation were not observed during the later sessions of Phase II. Instead, the frequency of letters attempted approached zero levels of responding.

One final comment should be made related to the choice of the independent variable selected for this study. One might question the selection of a frequency measure as the principal target behavior instead of requiring students to assess the accuracy with which they wrote the letter "m." It is acknowledged by the researcher that having subjects evaluate their behavior in terms of accuracy would be an appropriate target response. However, having subjects evaluate the accuracy of their attempts would have required more discrimination on the part of the subjects. Not only are finer and more discriminations required to evaluate accuracy of attempts, but having subjects evaluate accuracy alone would have meant that each attempt would have had to have been evaluated. Other than considerations for number and quality of discriminations, it seemed logical to first determine if self-evaluation procedures were powerful enough to even generate responding before targeting the quality of the responses made.
REFERENCES


BIBLIOGRAPHY


