BLACK, JOHN L. The Effect of Feedback on the Reliability of Self-Recording. (1974) Directed by: Dr. Rosemery O. Nelson. Pp. 84.

The purpose of this study was to determine if the reliability of self-recorders could be improved by providing different types of feedback contingent on the reliability of their self-recording. The feedback varied in amount of information: the non-specific feedback group received generalized social reinforcement in order to maintain their self-recording, the verbal feedback group received verbal feedback contingent on reliability, and the verbal plus numerical feedback group was given verbal and numerical feedback, also contingent on reliability. All three groups (N = 7 college student subjects per group) were observed under four conditions. The first was a baseline period during which independent observers took observations of the subjects who were unaware of the observers and unaware that they would be asked to participate in the study. During the second condition the subjects self-recorded and were aware that an independent observer was assessing the reliability of their recording. This condition was interspersed with a condition in which the subjects continued to self-record but were unaware that reliability was being assessed. The final condition was a return-to-baseline, in which the subjects were told the study was over and did not self-record; the independent observers, however, continued recording data. The target

behavior was face-touching, and the study occurred in a classroom situation.

Self-recording was found to be reactive. The frequency of face-touching decreased significantly during the self-recording conditions, but showed a significant increase in frequency during return-to-baseline.

An analysis of variance found no significant difference in reliability among the three treatment groups. Also, no significant difference was found in reliability between the conditions in which the subjects were aware and unaware of reliability checks.

THE EFFECT OF FEEDBACK ON THE RELIABILITY OF SELF-RECORDING

by

John L. Black

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
1974

Approved by

Rosemary O. Nelson
Thesis Adviser

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of the Graduate School at The University of North Carolina at Greensboro.

Thesis Adviser Rosemary O. Nelson

Committee Members

September 17, 1974
Date of Acceptance by Committee

ACKNOWLEDGMENTS

I would like to thank Dr. Rosemery Nelson, my thesis adviser, for the ideas and assistance she provided in carrying out and writing up this study. I would also like to express my appreciation to Dr. Robert Jones and Dr. Jacquelyn Gaebelein for their advice and assistance as my committee members. Special thanks are extended to Dr. Scott Lawrence, whose class served as subjects for this study, and to Dr. David Lipinski and Mary Newell, who assisted in data collection.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Self-Recording

Rationale

The most frequently used assessment technique in behavior modification is direct observation of the target behavior in the natural environment. Assessment is done in the natural environment because that is where the behavior is occurring, and in order to determine the antecedents and consequents of the behavior. However, assessment in the natural environment is very impractical if an independent observer is required to be present to observe and record the target behavior. This problem is circumvented by using instruments (Azrin & Powell, 1968), or the subject to record the behavior (Goldiamond, 1965; Stuart, 1967). Recording with instrumentation yields reliable data, but it is bulky, expensive, fragile, and impractical in many situations. Thus, self-recording of behavior is advisable for convenience reasons. It is also the only means of assessment for covert behaviors without public referrants, e.g., urges to smoke (McFall & Hammen, 1971). In addition, self-recording is needed for assessing "private" behaviors, e.g., bed wetting and sexual behavior.

Although self-recording appears to have advantages as an assessment device, there are many problems inherent

in its use. The main problems are reactivity and lack of reliability of self-recording.

Reactivity

Reactivity refers to the phenomenon that the target behavior is altered as a result of observation. Independent observers have been found to have a reactive effect on the behavior of nursery school children (Arsenian, 1943), delinquent children (Polansky, Freeman, Horowitz, Irwin, Papanis, Rappaport, & Whaley, 1949), visitors to an art museum (Bechtel, 1967), and families in the natural environment (Patterson & Harris, 1968).

The self-recording of behavior also has reactive effects on the target behavior. The reactivity of self-recording seems to change behavior in therapeutic directions. Self-recording the frequency of inappropriate behaviors seems to lead to a reduced frequency of the behaviors, while self-recording appropriate behaviors seems to produce an increase in frequency. Inappropriate behaviors which have been reduced merely by recording their frequency are facial tics, out of seat, inappropriate hand raising in class, repetitive scratching, fingernail biting (Maletzky, 1974), talking out in class (Broden, Hall, & Mitts, 1971), reported hallucinations (Rutner & Bugle, 1969), excessive eating (Mahoney, 1974), and face-touching (Lipinski & Nelson, 1974). Appropriate behaviors which

have been increased by self-recording are study behavior (Johnson & White, 1971; Broden et al., 1971), oral class participation (Gottman & McFall, 1972), and resisting urges to smoke (McFall, 1970; McFall & Hammen, 1971).

Although it appears that self-recording produces reactive effects in a therapeutic direction, there are problems in maintaining the behavior change. McFall (1970) and McFall and Hammen (1971) reported that a decrease in smoking produced by self-recording was maintained. Gottman and McFall (1972) found that the increase in in-class participation was maintained in follow-up. However, several other studies have found the reactive effects of self-recording to be short-term. Maletzky (1974) reported decreases in fingernail biting, repetitive scratching, inappropriate hand raising in class, facial tics, and out of seat by having the subjects self-record these behaviors on a wrist counter. Although the frequency of the behaviors was greatly reduced when the counter was worn, the frequency of the behaviors quickly increased when the recording procedure was discontinued. Long term effects were achieved by gradually reducing the use of the counter.

Lipinski and Nelson (1974) also found the reactive effects of self-recording to be short-term. The frequency of face-touching by college students in a class was significantly reduced by self-recording; but when recording was

discontinued, the frequency quickly returned to the base rate. This short-term reactive effect on face-touching has been replicated (Nelson, Lipinski, & Black, in press).

The short-term reactive effects of self-recording have also been found in a weight control study (Mahoney, 1974). For the first two weeks of the program, three treatment groups (self-reward for weight loss, self-reward for eating habit improvement, and self-recording only) self-recorded their weight and eating habits which resulted in a significant loss of weight for all three groups. For the following six weeks, the self-recording group continued to self-record and, in addition, set weekly weight loss and habit improvement goals; but there was no significant loss in weight. After the eight-week study was over, the self-reward for weight loss and the self-reward for habit improvement groups were found to be superior to a control group, whereas there was no difference between the self-recorders and controls. Although the self-recorders did not relapse and gain additional weight, the reactive effects dissipated after only two weeks.

A final study finding short-term reactive effects of self-recording was done by Broden et al. (1971). Studying had been increased in one student and talking out had been decreased in another student by self-recording, but there was a return to the base rate for both behaviors when self-recording was discontinued.

A factor which may play a part in the maintenance of behavior changed by self-recording is the stimulus properties of the recording apparatus. The studies of Broden et al. (1971) and Maletzky (1974) indicate that the recording apparatus may serve as a discriminative stimulus to produce or inhibit the target behavior. In the Broden et al. (1971) study, an eighth-grade girl increased her in-class study behavior by self-recording the target behavior. Withdrawal of the data slips which she used for self-recording resulted in a decrease in study behavior; whereas reinstituting the slips increased the behavior. It appears that the slips functioned as a discriminative stimulus for studying, instead of as a mere necessity for self-recording. Study behavior decreased on days when slips were not issued, although on previous days the student had studied even though she had been forgetting to record.

In the study by Maletzky (1974), self-recording was used to decrease inappropriate behaviors. The importance of the recording apparatus, a wrist counter, as a discriminative stimulus can be supported by the data. When self-recording was terminated, the frequency of the behaviors increased, but it cannot be determined if this was due solely to the removal of the discriminative stimulus or the termination of recording or both. The verbal reports of the subjects indicate that the discriminative stimulus

function of the apparatus played some part. "Patients typically noted the discriminative stimulus quality of wearing the counter on their wrists; as Case 1 commented 'I'm more aware of not scratching when I feel that thing on my wrist'" (Maletzky, 1974, p. 110). In the Broden et al. (1971) study, generalization was facilitated by combining self-recording with teacher praise, followed by teacher praise alone, and then a return-to-baseline. Maletzky (1974) generalized reactivity by gradually decreasing the use of the wrist counters.

In addition to the stimulus function of the recording apparatus, other factors which may influence the maintenance of behavior change produced by self-recording are: the subject's motivation, the specific target behavior being recorded, the duration of self-recording, and the value orientation of the behavior (Lipinski, 1974). With regard to the subject's motivation, McFall and Hammen (1971) reported that their subjects wanted to stop smoking and that the decrease in smoking was maintained in follow-up. The reactive effects of self-recording were short-term in the studies by Broden et al. (1971), Lipinski and Nelson (1974), and Nelson et al. (in press). Unlike McFall and Hammen's subjects, these latter subjects did not indicate that they wished to change their behavior.

Another factor which may determine if the behavior change is long- or short-term is the nature of the target behavior. The behavior change has been maintained when the target behavior was smoking (McFall, 1970; McFall & Hammen, 1971), oral class participation (Gottman & McFall, 1972), and weight loss (Mahoney, 1974). The behavior change was not maintained when the target behavior was face-touching (Lipinski & Nelson, 1974; Nelson et al., in press), talking out, and studying (Broden et al., 1971). If the behavior change is valued by others, the change may be maintained by social reinforcement, e.g., loss of weight. Even if the behavior change is not valued by others, it could be maintained by self-reinforcement. If the subject placed a high value on the change in behavior, the "feedback loop" hypothesized by Kanfer (1970) may be operating. When the subject self-records, he is attending to the behavior. This feedback from self-observation is compared with the desired performance criteria; if the behavior observed is at the criterion level, the subject engages in self-reinforcement. Thus, the behavior change took place because the high value judgment of the behavior resulted in self-adjustive behaviors. The behavior change could also be maintained if the new behavior was intrinsically reinforcing, e.g., developing athletic skills.

The duration of self-recording also may influence the maintenance of behavior change. A stable behavior with a long history of reinforcement would require a longer self-recording period. If the self-recording period was too short, self-control would not be established, and the behavior would return to the base rate upon termination of self-recording (Lipinski, 1974).

In addition to the maintenance of behavior change, a second problem with the reactivity of self-recording is controlling the direction of reactivity. It appears that the self-recording of appropriate behaviors results in an increase in the frequency of these behaviors whereas recording inappropriate behaviors produces a decrease in the frequency of these latter behaviors. Thus, one factor which may influence the direction of reactivity is the value judgment of the target behavior. Research on smoking has been relevant to the problem of direction of reactivity because the frequency of smoking has been reduced by self-recording when smoking has been labeled as inappropriate. McFall and Hammen (1971) used four self-recording groups. One group recorded the number of cigarettes smoked per day. A second group recorded the number of times they were unable to resist the urge to smoke. A third group recorded their frequency of resisting the urge to smoke. A fourth group was required to successfully resist the urge to smoke

20 times per day and to record this. All four groups showed a drop in smoking rate, but there were no significant differences between groups. When smoking was labeled as an appropriate behavior, e.g., by means of the professor modeling at the beginning of class, self-recording increased the in-class frequency of smoking (McFall, 1970).

Nelson, Lipinski, and Black (1974) also investigated the effect of value judgment on the direction of reactivity of self-recording. The subjects in the study were adult retardates, who were assigned to one of three groups. One group self-recorded the frequency of face-touching, which the experimenters labeled as undesirable; a second group self-recorded the frequency of talking, which was labeled as desirable; and a third group self-recorded the frequency of object-touching, which was considered to be a neutral behavior. Self-recording resulted in an increase in the frequency of the desirable behavior (talking), and produced decreases in the frequency of the undesirable behavior (face-touching) and of the neutral behavior (object-touching). When the subjects were reinforced for reliable recording, talking and object-touching increased in frequency and face-touching decreased in frequency.

The studies of McFall (1970), McFall and Hammen (1971), and Nelson et al. (1974) indicate that labeling a behavior as appropriate or inappropriate may affect the direction of

reactivity. Other factors which may affect the direction of reactivity that have been investigated are instructions and expectancy. Gottman and McFall (1972) looked at the influence of instructions using a crossover design. One group of subjects self-recorded oral class participation while the other group self-recorded nonparticipation. The two groups were next instructed to reverse their self-recording, i.e., the group recording oral participation began recording nonparticipation and vice versa. The reactive effects of self-recording were found to be influenced by instructions. Both groups had a significant increase in oral class participation when recording this behavior, but their talking decreased when they recorded nonparticipation.

In a study to determine the effect of expectancy on the direction of reactivity, four groups of subjects were given differential expectancies of how self-recording would affect the frequency of face-touching: increase, decrease, no change, and no expectancy (Nelson et al., in press). Subjects self-recorded the frequency of their face-touching which resulted in a decrease in the behavior in all four groups. Unlike instructions, expectancy appears to be ineffective in altering the direction of reactivity.

Although the reactivity of self-recording is beneficial in that the behavior changes produced are usually in a therapeutic direction, there is a problem in differentially

analyzing the reactive effects of self-recording and the effects of other therapeutic components. In many studies using self-recording in a therapeutic program, self-recording has been combined with other self-control procedures. As a result, it cannot be determined if the behavior change is due to the reactivity of self-recording or to other self-control procedures. Nelson and McReynolds (1971) have proposed that this problem may be overcome by appropriate experimental design. A control group which engages only in self-recording can be compared with experimental groups which engage in self-recording and other self-control techniques.

Reliability

A second major problem with self-recording is reliability. The reliability problem has two components. Whether or not it is necessary to assess the reliability of the self-recorders; and, when reliability is assessed, the fact that reliability is low.

The question of whether or not reliability must be assessed has been viewed from three positions. According to Simkins (1971a, 1971b), assessment of interobserver reliability is necessary in self-recording studies in order to determine the effectiveness of the therapeutic techniques employed. In contrast, Watson and Tharp (1972) do not

consider assessment of reliability to be a necessity. They view the reliability problem as persuading the subject to consistently engage in recording behavior when the target behavior occurs rather than waiting and depending on memory. If the behavior is not consistently recorded, the advantage of reactive effects on the target behavior is lost, as well as there being no data on the target behavior. A third position on the question is that of Nelson and McReynolds (1971). They differentiate between the reliability of self-recording and the reliability of the effects of self-recording. It may be possible for the recording to be unreliable but for the behavior change due to reactivity to be reliable. Simkins' view reflects an emphasis on purity in research, whereas Watson and Tharp and Nelson and McReynolds stress applicability in the natural environment.

Whichever position on the question of reliability assessment is adopted, the data indicate that self-recorders are unreliable as compared with independent observers. McFall (1970) reported a correlation of .61 between self-recorders and independent observers. In this study, the subjects were unaware that their reliability was being assessed. The reliability of self-recording has been found to increase when recorders are aware that reliability is being assessed. Lipinski and Nelson (1974) report a reliability between self-recorders and independent observers of

.86 when they were aware that they were being observed, and .52 when they were unaware of reliability assessment. This result has been replicated by Nelson et al. (in press). The reliability coefficient was .810 when self-recorders were aware that their reliability was being assessed, and .554 when they were unaware of the independent observer's assessment. This decrease in the reliability of self-recorders when they are unaware that their reliability is being assessed is comparable to independent observers. Reid (1971) found that the median reliability of observers was .75 when they were aware of reliability assessment, but reliability dropped to .51 when they thought reliability was not being assessed. Taplin and Reid (1973) found that the reliability of observers was .81 on the last day of observer training but that reliability decreased to .65 on the first day observers were unaware of reliability checks. Romanczyk, Kent, Diament, and O'Leary (1973) also report that observer reliability was lower when they were unaware of reliability checks than when they were aware of them.

Although self-recorders have been found to be unreliable, recent research has indicated that the reliability of self-recorders can be increased by training and/or reinforcement. Flowers (1972) reduced the cheating behavior of a sixth-grade girl by reinforcing reliable self-evaluation. The student was required to grade her own assignments on

Monday through Thursday. On Friday the teacher gave her a test which was based on material from the previous assignments. The student's grade was based on reliable self-evaluation, i.e., the Friday test score was compared with the mean performance for Monday through Thursday and the greater the discrepancy between the two scores the lower the grade. The frequency of cheating dropped to zero and the student's grades rose significantly following treatment. In another study using elementary school children, Bolstad and Johnson (1972) increased the reliability of self-recording by using reinforcement. The subjects were required to record disruptive behaviors. If the subjects recorded low frequencies of disruptive behavior and if their recordings were within limits of the observer's recordings, the subjects were given points which were exchangeable for prizes.

A procedure similar to that of Bolstad and Johnson (1972) was used to increase the reliability of self-report of boys in a community-based behavior modification program for delinquents (Fixsen, Phillips, & Wolf, 1972). The boys were trained on the behavioral definitions of the target behaviors and then were reinforced by making points in a token economy contingent upon agreement between self-report and peer-reports.

The reliability of self-recording by college students has also been improved by means of reinforcement. Lipinski (1974) differentially reinforced college students either for increasing the reliability of their self-recording or for decreasing the frequency of the target behavior. The reliability of the group reinforced for increased reliability was significantly better ($p < .05$) than the reliability of the group reinforced for decreasing the frequency of the target behavior.

In a final study, the reliability of the self-recording of adult retardates was increased by reinforcement (Nelson et al., 1974). There were three groups of subjects: one group self-recorded face-touching; one group self-recorded object-touching; and one group self-recorded talking. Reinforcement resulted in a significant increase in the reliability of the subjects self-recording face-touching and object-touching but the reliability of the subjects recording talking remained unchanged.

In summary, it has been found that self-recording can be used as a means of assessment and as a treatment procedure. As a treatment procedure it is recommended due to its having reactive effects in a therapeutic direction. As a means of assessment it is recommended because it is convenient compared with assessment using independent observers or instrumentation, and it is the only means of assessment

when the target behavior is covert or private. Self-recording has the disadvantage of being reliable only when the self-observer knows that an independent observer is assessing the reliability of his recording. In order that self-recording can be used as a more accurate means of assessment, the reliability must be improved. Improving reliability may also make self-recording a more valuable treatment procedure by making Kanfer's feedback loop (self-observation, self-evaluation, and self-reinforcement) more effective and thus producing greater reactivity. Although some factors influencing the reactivity of self-recording have been investigated, with the exception of training and reinforcement, the variables influencing the reliability of self-recording have not. One possible factor which may increase the reliability of self-recording is feedback. The purpose of this study was to determine if the reliability of self-recording could be improved by providing different types of feedback contingent on the reliability of self-recording. The following section reviews the effect of feedback on performance.

Feedback

In the following pages, the types of feedback, forms of feedback, tasks, and subjects used in the feedback studies will be described. The basic studies, results of

these studies, explanations for the results, and experiments testing the theoretical explanations will be included.

Types of Feedback

Positive feedback, negative feedback, and combinations of these two have been used. The basic combinations of feedback used have been those devised by Buss and his associates (Buss & Buss, 1956; Buss, Braden, Orgel, & Buss, 1956). The basic feedback combinations are: Right-Wrong, Right-Nothing, and Nothing-Wrong. In the Right-Wrong feedback condition, the subject is told "right" for correct responses and "wrong" for incorrect responses. In the Right-Nothing condition the subject is told "right" for correct responses and is given no feedback for errors. In the Nothing-Wrong condition the subject is given no feedback for correct responses and told "wrong" for incorrect responses. The Right-Wrong condition provides feedback 100% of the time whereas the other two conditions initially provide feedback only 50% of this time. As learning occurs this percentage increases in the Right-Nothing feedback condition and decreases in the Nothing-Wrong condition.

Tasks

The most frequent task employed in studies investigating the effects of various types of feedback have been discrimination learning. In addition to discrimination

tasks, the influence of feedback on concept learning (Cairns, 1967; Siegel & Downey, 1970), academic response rate (Walker & Buckley, 1972), student achievement (Lovett, 1972), and motor performance (Talkington, Altman, & Grinnell, 1971) have also been investigated.

Populations Studied

The form of feedback has differed in studying animal populations, since verbal feedback would be inappropriate. In animal studies involving discrimination learning, food was provided for correct responses and shock for errors. The result was a facilitation in learning (Hoge & Stockling, 1912; Warden & Aylesworth, 1927).

The research with human subjects has employed several different populations: elementary school children (Cairns, 1967; Meyer & Offenbach, 1962; Penney & Lupton, 1961; Spear, 1970; Walker & Buckley, 1972; Willcutt & Kennedy, 1963), kindergarten age children (Brackbill & O'Hara, 1958; Terrell & Kennedy, 1957), mongoloids (Talkington *et al.*, 1971), neuropsychiatric patients (Buss & Buss, 1956), and college students (Lair & Smith, 1970; Lovett, 1972; Mueller & Gumina, 1972; Siegel & Downey, 1970).

Basic Studies

This section will describe some of the earlier studies which used the three basic types of feedback and the results of these studies.

In a study with neuropsychiatric patients as subjects, Buss and Buss (1956) found the Right-Wrong and Nothing-Wrong conditions to be equivalent, and superior to Right-Nothing feedback in learning a color concept. In a second experiment, neuropsychiatric patients and student nurses learned a number concept and then a color concept. There was no difference between the Right-Wrong and Nothing-Wrong groups in learning either concept. Buss et al. (1956) again used neuropsychiatric patients in investigating the influence of Right-Wrong and Nothing-Wrong feedback on the acquisition and extinction of a discrimination. Again, the Right-Wrong and Nothing-Wrong conditions were equivalent, and superior to Right-Nothing feedback.

Kindergarten children were used by Brackbill and O'Hara (1958) in a discrimination task involving three boxes in which the correct response was the position alternation of the first two boxes. For correct responses the subjects were rewarded instead of being told "right" and were punished for errors instead of being told "wrong." If the word "right' is assumed to be a positive reinforcer and "wrong" is assumed to be a punisher, then the results are like those of the previous studies. A reward-punishment group was found to be superior to a reward-nothing group.

Meyer and Offenbach (1962) compared the influence of the three basic feedback conditions on the learning of a

block discrimination task by third- and fourth-grade students. The task had three levels of complexity and when there were two or three irrelevant dimensions in the task, the Nothing-Wrong and Right-Wrong conditions were equally effective, and both were superior to the Right-Nothing condition. However, with only one irrelevant dimension there were no significant differences between the three groups.

Another study with the three basic feedback conditions had mongoloid subjects learn a simple alternation of dropping marbles in two holes. Talkington et al. (1971) found the Nothing-Wrong group to be significantly better than the other two groups ($p < .001$) and the Right-Wrong group to be significantly better than the Right-Nothing group ($p < .05$).

In an experiment by Penney and Lupton (1961), subjects were rewarded with jelly beans for correct responses and punished with an intense tone for errors. The experimenters used conditions of Reward-Nothing, Reward-Punishment and Nothing-Punishment with students from grades two, four, and eight. The Nothing-Punishment group learned the visual discrimination task faster than the Punishment-Reward group which learned it faster than the Reward-Nothing group. These results are similar to the Right-Wrong feedback studies if "right" is assumed to be a reinforcer and "wrong" is assumed to be a punisher.

Explanation of Results

Reinforcement Explanation. The first explanation for the superiority of Nothing-Wrong and Right-Wrong feedback over Right-Nothing feedback was offered by Buss et al. (1956) in terms of reinforcement. No feedback was assumed to provide no reinforcement and wrong feedback was considered to be a stronger negative reinforcer than right feedback as a positive reinforcer. A main problem in this explanation is the misuse of terminology. Buss et al. are labeling wrong feedback as negative reinforcement instead of punishment. If wrong feedback acted as a negative reinforcer then the frequency of errors should increase, but since the frequency of errors decreases, wrong feedback must be acting as a punisher. This explanation can also be interpreted as indicating that escape from the aversive consequence of being told "wrong" is a stronger negative reinforcer than being told "right" is a positive reinforcer.

Motivation Explanation. Brackbill and O'Hara (1958) explained the superiority of a Reward-Punishment group over a Reward-Nothing group in learning a discrimination task in terms of motivation. They stated that learning is faster in a high drive condition than in a low drive condition. Both punishment and reward increase drive, therefore, a Punishment-Reward condition would produce higher drive than

the Reward-Nothing condition, which would result in faster learning in the Reward-Punishment group.

Meaning Induction Explanation. Another explanation, the meaning induction hypothesis, was presented by Buchwald (1959a, 1959b). Buchwald said that although no feedback originally provides no information, it eventually represents the type of feedback opposite of the overt feedback given to the subject. For example, in the Nothing-Wrong condition no feedback comes to mean "right" rather than remaining at a no information level and in the Right-Nothing condition no feedback comes to mean "wrong." The Nothing-Wrong condition is superior to the Right-Nothing condition because meaning induction is supposedly faster in the Nothing-Wrong condition.

Information Processing Explanation. The information processing hypothesis explains the difference in the feedback conditions as due to the Right-Nothing feedback providing more ambiguous information than the Nothing-Wrong feedback (Buchwald, 1962). When the subject is told "wrong" he knows that both his particular response and his hypothesis are wrong but when he is told "right" he knows only the response is correct but not if the hypothesis is correct. For example, if the subject is presented with a stimulus which is not an example of the correct concept and he says

"no," the experimenter will tell him "right." This provides no information as to whether his hypothesis should be changed because there can be many negative examples of the correct concept. However, if the subject is presented with a negative example of the concept and says "yes" the experimenter will tell him "no." This informs him that his hypothesis was wrong and that it must be changed.

Studies Testing the Explanations for Feedback Results

The reinforcement explanation states that wrong feedback is a more effective punisher than right is as a reinforcer. The information processing hypothesis explains the superiority of Right-Wrong and Nothing-Wrong feedback over Right-Nothing feedback as due to less ambiguous information in the wrong than in the right feedback. A study by Mueller and Gumina (1972) compared right and wrong feedback by controlling the value of the nothing feedback. By controlling the value of the nothing feedback, the study also tested the meaning induction hypothesis. The three basic feedback conditions were used in a two-choice concept discrimination task. The value of nothing feedback was controlled by giving half of the subjects explicit instructions to attribute no information to the nothing feedback. The other half were given standard instructions, which made no mention of the nothing feedback. Therefore, the subjects in this group could engage in meaning induction.

Both the reinforcement and information processing explanations would predict that Nothing-Wrong condition would be superior to Right-Nothing and this superiority would be more pronounced with explicit instructions. There was little difference between the Right-Nothing and Nothing-Wrong conditions under explicit instructions, therefore, it appears wrong is not an inherently superior source of feedback. The Right-Wrong condition was found to be superior to the other two conditions under both standard and explicit instructions. Since the Right-Wrong group was given feedback 100% of the time whereas the other two received feedback 50% of the time, it appears that learning is facilitated by the frequency and amount of information in the feedback rather than wrong being superior to right.

A second experiment which manipulated the frequency of the types of feedback was done by Siegel and Downey (1970). College students learning a concept were given either Nothing-Wrong or Right-Nothing feedback. The authors hypothesized that previous results showing Nothing-Wrong feedback to be superior to Right-Nothing feedback were due to more frequent negative feedback. To control for this each group was given feedback at an average rate of one response in three. Rather than the groups being equal as predicted, the Nothing-Wrong group was significantly better ($p < .05$). These results can be explained

by the amount of information in the feedback instead of the reinforcement value of the feedback. The negative feedback provides information only on the particular response. If the subject is presented with a stimulus which is not an example of the concept and he responded "no" he would be told "right." This provides no information as to whether his hypothesis should be changed because there can be many negative examples of the correct concept.

Another study which indicates information is the key factor in feedback used second- and third-graders. Elovson (1971) simultaneously presented the subjects with four CVC trigrams and an English word. They were to select the trigram that was "synonymous" with the English word. The correctness of a response was determined by a previously-arranged reinforcement schedule. The five treatment groups were reward, penalty, confirmation, information, and penalty plus information. Subjects in the penalty group committed significantly more errors than either the information group or the penalty plus information group. It was not the presence of penalty that was the critical factor in making errors but the lack of information. It was also found that the subjects made significantly more correct responses following errors receiving information or penalty plus information feedback than following correct responses which were positively reinforced. Thus, learning was

facilitated by informative feedback, not positive or negative feedback.

A final study indicating information is the critical factor in feedback used academic response rate as the dependent measure. Two students in a class for behaviorally disordered children were used in this study by Walker and Buckley (1972). The task was simple division problems. The experimental subject received Right-Wrong feedback in the first phase of treatment, positive reinforcement for correct responses and punishment for errors in the second phase, and in the third phase received Right feedback plus positive reinforcement for correct responses and Wrong feedback plus punishment for incorrect responses. The other student served as a control. During baseline the experimental subject's error rate was twice the correct rate. In the Right-Wrong feedback phase the error rate was only slightly more than the correct rate. The error rate again was only slightly more than the correct rate in the positive reinforcement-punishment phase, but both error and correct rates were almost double those of the Right-Wrong feedback phase. In the third phase, in which the subject received positive feedback and positive reinforcement for correct responses and negative feedback and punishment for errors, the overall response rate was the same as in phase two but the correct rate was triple the error rate. Thus,

the more feedback provided, the better the learning and performance when the dependent measure is academic response rate.

These studies have cast doubt on any explanations for the superiority of Right-Wrong and Nothing-Wrong feedback to Right-Nothing feedback based on the inherent superiority of negative feedback over positive feedback. There is evidence indicating that the key factor in the effect of feedback on learning and performance is the amount of information in the feedback.

Form of Feedback

Most of the feedback studies have given the feedback in verbal form. Cairns (1967) used fourth-grade students in a study in which they received feedback either verbally or by means of a buzzer. The verbal group was told "right" for correct responses and nothing for errors. The nonverbal group heard the buzzer for correct responses and nothing for errors. Half of each group were informed, i.e., they were told the buzzer or "right" meant the response was correct and the other half of each group received no information regarding the meaning of feedback. A very strong information effect was found. Both the verbal and nonverbal group performed significantly better ($p < .001$) when they knew what the feedback meant.

Lair and Smith (1970) performed a very similar experiment in which subjects received verbal feedback ("right" for correct responses and nothing for errors) or non-verbal feedback (red light for correct responses and nothing for errors). Subjects were also informed or not informed as to the meaning of the feedback. Like Cairns' study, a discrimination task was used but the subjects were college students. In contrast to Cairns' results they found a significant verbal feedback effect ($p < .01$). The college students receiving verbal feedback performed better than those receiving non-verbal feedback in both the informed and uninformed conditions. It can be concluded that perhaps feedback in verbal form is more effective than non-verbal feedback when using college students as subjects.

Statement of the Problem

Examination of research in self-recording indicates that self-recording can be used as a means of assessment and as a treatment procedure. As a means of assessment it is recommended because it is convenient compared with assessment using independent observers or instrumentation, and it is the only means of assessment when the target behavior is covert or private. Although self-recording has advantages as an assessment device, there are many problems inherent in this use. In addition to reactivity, a main

problem is the unreliability of self-recorders compared with independent observers. In order for self-recording to be used as an accurate means of data collection, the reliability must be improved. At present the only factors influencing the reliability of self-recording which have been investigated have been reinforcement and training. Another factor which may influence the reliability of self-recording is feedback. Examination of the feedback literature has shown that the key factor in the influence of feedback on learning and performance is the amount of information in the feedback. The purpose of this study was to determine if the reliability of self-recorders could be improved by providing different types of feedback contingent on the reliability of their self-recording. The feedback varied in the amount of information since this is the crucial factor. One group received generalized social reinforcement in order to maintain their self-recording, a second group received verbal feedback contingent on reliability, and a third group was given verbal and numerical feedback, also contingent on reliability. All three groups were observed under four conditions. The first was a baseline period during which independent observers took observations of the subjects who were unaware of the observers and unaware that they would be asked to participate in the study. During the second condition the subjects

self-recorded and were aware that an independent observer was assessing the reliability of their recording. This condition was interspersed with a condition in which the subjects continued to self-record but were unaware that reliability was being assessed. The final condition was a return-to-baseline, in which the subjects were told the study was over and did not self-record; the independent observers, however, continued recording data.

Based on the feedback literature, a prediction can be made that the verbal plus numerical group should be the most reliable, since this feedback would provide the most information. The verbal feedback group should be less reliable, since this provides less information, and the non-specific feedback group should be the least reliable in self-recording.

CHAPTER II

METHOD

Experimental Design

The experimental design was a 3x4x7 factorial design with subjects nested in the three treatment groups and repeated across the four experimental conditions with seven observations under each condition.

Subjects

The subjects were 21 college students (three males, 18 females) who were in a graduate class for continuing education in psychology which met for an hour and a half two times per week for a semester. The subjects were equally divided into three groups and were matched in the groups. Matching was accomplished by using the following procedure. First, the subjects were rank ordered on the basis of the mean frequency of the target behavior during baseline. The subjects were then assigned to the groups as follows: the subject with the highest mean frequency was assigned to the first group; the subject with the next highest mean frequency was assigned to the second group; and the subject with the third highest frequency was assigned to the third group. The subject with the fourth highest mean frequency was assigned to the third group and the subjects with the fifth and sixth highest mean frequencies were assigned to the second and first groups

respectively. This procedure was followed until all 21 subjects were equally assigned to the three groups.

Independent Observers

Three independent observers (two males, one female) were used during all conditions. During the baseline, self-recording unaware, and return-to-baseline conditions, the observers remained unobtrusive by observing from behind a one-way mirror in a room adjacent to the class. During the self-recording aware condition, one observer was present in the classroom while the other two remained behind the one-way mirror. The independent observers assessed the reliability of the self-recorders as well as their own reliability.

Target Behavior

An overt target behavior was used in order that inter-observer reliability could be assessed, which would not be possible with a covert behavior. A second criteria for the target behavior was a high in-class frequency. The target behavior was face-touching which was defined as touching any part of the body from the neck up with the hand or an object held in the hand. For a new behavior to occur, the hand or object had to break contact with the face and then return to the face. If a subject moved his hand over different parts of the face while maintaining contact, it was

counted as only one behavior. If both hands touched the face simultaneously at different places, it was counted as two behaviors. Objects held in the hand were considered part of the hand; and objects in contact with the face, e.g., glasses, were considered part of the face.

Conditions

Baseline (Base). During baseline, the subject's frequency of face-touching was recorded by the independent observers from behind the one-way mirror. The subjects had no knowledge of the presence of the observers behind the one-way mirror nor did they know that they would be asked to participate in the study. Baseline continued until seven data points were taken on each subject. Each data point consisted of the frequency of face-touching per 5-minute interval.

Self-Recording (S-R Aware and S-R Unaware). Prior to the initiation of the self-recording conditions, subjects were selected and trained. Subjects were selected on the basis of a high frequency of face-touching during baseline. Those subjects with a high frequency of face-touching were recruited by being offered a cash payment which would not be less than three dollars. At the completion of the study, each subject was paid four dollars.

The training procedure was as follows:

(1) The subjects were given a data sheet which listed the rules for self-recording and defined face-touching. The use of the data sheets in recording and the role of the independent observer in assessing the reliability of their recording was described.

(2) A model engaged in face-touching while one of the experimenters recorded the behavior and explained to the subjects why a particular behavior was or was not a face-touch.

(3) The model engaged in face-touching and the subjects recorded the frequency of the behavior.

(4) In the final stage the subjects self-recorded the frequency of their face-touching during a 2-minute interval.

In order to reach criterion, the subjects had to correctly record the frequency of the model's behavior and then practice recording their own face-touching for a 2-minute interval. After recording the model's behavior, all subjects self-recorded for a 2-minute interval. Those who had recorded the model's behavior correctly then left. Those who had incorrectly recorded the model's behavior remained and recorded the face-touching of the model again, and then practiced self-recording for another 2-minute interval. The face-touches of the model served as a standard criterion instead of the face-touches of each subject since it was a group training procedure.

On the first day of self-recording, the subjects were given a sheet of written instructions, a copy of which is included in Appendix A. The subjects were instructed to fill out the data sheets completely. When they forgot to self-record for an interval, they were to write "forgot" rather than zero, which meant no face-touches. The subjects also were informed that they would be receiving written feedback on their self-recording and were asked not to discuss their feedback with other members of the class.

During the self-recording condition in which subjects were aware that the reliability of their self-recording was being assessed (S-R Aware), one of the independent observers was in the classroom and the other two independent observers recorded from behind the one-way mirror. Seven data points were collected per subject during this condition. Each data point consisted of the frequency of face-touching per 5-minute interval.

In the other self-recording condition, S-R Unaware, the subjects continued to self-record but the independent observers were behind the one-way mirror with the subjects unaware of their presence. Seven data points were again collected on each subject.

Return-to-Baseline (RT-Base). The subjects were told that the study was over, but in fact the independent

observers continued recording and gathered seven data points on each subject. As a check on the success of the unobtrusiveness of the observers during all but the S-R Aware condition, a post-experimental questionnaire was given to all subjects. The questionnaire contained among others, the question, "What aspects of this study, if any, aroused your suspicion?" A copy of the questionnaire is included in Appendix B.

Treatment Groups

There were three treatment groups with the levels of the treatment factor being the amount of information in the feedback. The non-specific feedback group received no information concerning the reliability of their self-recording. The verbal feedback group was given verbal information on the reliability of their self-recording. The verbal plus numerical feedback group was given both verbal and numerical information on the reliability of their recording. All feedback was provided in written form, attached to the subject's blank data sheets. The data sheets and feedback were either handed out or placed on the subject's desk by one of the independent observers prior to each class. The basis for the feedback was the subject's self-recording during the S-R Aware condition for the last class the subject attended. A right-wrong procedure was

used, i.e., the subjects received positive feedback for increased reliability and negative feedback for decreased reliability. Samples of the forms on which the subjects received feedback are included in Appendix C.

Non-Specific Feedback (N-S F). This was a control group which received no feedback on the reliability of their recording. They were reinforced for filling out data sheets completely, and for turning them in regularly; or they were prompted if they did not do so. Subjects in this group received non-specific feedback on form one, which is included in Appendix C.

Verbal Feedback (VF). The subjects in this group were given verbal feedback as to the reliability of their self-recording. They received a verbal statement which was determined by the per cent of change in their reliability from the last class they attended. When the subjects reached a high degree of reliability, .90, they were then given a maintenance statement. These subjects received their feedback of form two in Appendix C. Each verbal statement is matched with the corresponding change in reliability in Table 1. All tables and figures are located in Appendix D.

Verbal plus Numerical Feedback (V+NF). This group received verbal feedback contingent on their reliability in the same manner that the VF group did. In addition to the verbal feedback, they were informed as to their calculated reliability and the change in reliability from the previous session in which their reliability was assessed. The subjects received their feedback on form three, which is included in Appendix C.

If any subject turned in an incomplete data sheet the subject was given form four, which was a prompt to fill out data sheets completely. In the case of absence, the subject was given form five upon returning to class. This was a prompt to fill in all intervals on the data sheet on that day in order that the study could be completed before the end of the semester. Forms four and five are included in Appendix C.

Apparatus

The independent observers used stopwatches and hand counters when they were observing from behind the one-way mirror. During the S-R Aware condition, the independent observer in the classroom used the clock in the classroom instead of a stopwatch, and recorded the frequency of the target behavior by means of a tally on a data sheet instead of by the hand counter. Copies of the data sheets used by

the subjects and the independent observers are provided in Appendices D and E respectively. The subjects used data sheets consisting of 18 5-minute intervals. Each interval was labeled as to time, e.g., 4:00 - 4:04, 4:05 - 4:09, 4:10 - 4:14, etc.

Reliability

The reliability of the subjects' self-recording was assessed both when they were aware and unaware that reliability was being taken. The subjects received feedback on their reliability based on their recording during the S-R Aware condition only. Reliability between the independent observers and the subjects was assessed by having both the independent observer and the self-recorder count the frequency of face-touching during the same 5-minute interval. The subjects' reliability with the independent observer for feedback purposes was calculated as follows:

$$\frac{\text{lower frequency count of face-touching}}{\text{higher frequency count of face-touching}} = \% \text{ reliability}$$

The lower frequency count of face-touches by either the independent observer or the self-recorder was used in the numerator of the formula, and the frequency count of face-touches that was higher was placed in the denominator of the formula. If the reliability of the independent observers was also being assessed, a mean for the frequency

counts of the independent observers was computed and compared with the self-recorder's frequency count.

The reliability between two independent observers was assessed by having both observers counting the frequency of face-touching of the same subject during the same 5-minute interval. The reliability of the independent observers was assessed at least 20 per cent of the time during each of the four conditions. One of the observers was designated as the reliability checker; the other two observers were not aware of when he was assessing their reliability.

CHAPTER III

RESULTS

Reactivity

The mean number of face-touches recorded by the independent observers for each session was calculated for each of the three treatment groups. The means for the three groups for each session under each of the four experimental conditions are presented in Figure 1.

Examination of Figure 1 indicates there was a substantial drop in the frequency of face-touching from baseline to the self-recording conditions. There was an increase in face-touching at the end of the self-recording conditions. The reactivity of self-recording can also be seen in the mean face-touching frequency of each subject under each of the four conditions (Table 2).

The analysis of variance performed on the face-touching data of the independent observers, which is presented in Table 3, also indicates a difference between the experimental conditions (B) ($F = 74.9122$; $df = 3, 54$; $p < .01$). To assess the magnitude of the experimental effect, a utility index was calculated, $UI_B = .33$ (Dodd & Schultz, 1973). A post-hoc comparison of means using the Newman-Keuls test found the mean frequency during the Base condition to be significantly different ($p < .01$) from the frequency during the S-R Aware and S-R Unaware conditions

(Table 4). The RT-Base condition also differed significantly from the two self-recording conditions. There was no significant difference between the S-R Aware and the S-R Unaware conditions, nor was there a difference between the Base and RT-Base conditions.

There was no significant difference between the treatment groups, as was expected since the subjects were matched in groups on the basis of face-touching frequency during the Base condition.

In order to complete this analysis of variance, the RT-Base data for one subject in the verbal plus numerical group who had dropped out of school was estimated using the method described by Kirk (1968).

Reliability for Self-Recorders

The reliability of the self-recorders was calculated for both the S-R Aware and S-R Unaware conditions. When the reliability of the independent observers was being assessed, a mean of the frequency counts of the independent observers was calculated and compared with the self-recorder's frequency count. A Kendall correlation coefficient was calculated for each subject under both of the self-recording conditions, and an overall correlation was calculated for each of the three treatment groups under both conditions (Table 5). With ordinal data, either the

Kendall or Spearman correlation coefficients are appropriate; the Pearson coefficient is not appropriate, however, because it should be used only with interval scales (Nie, Bent, & Hull, 1970). The Kendall was chosen over the Spearman for the present data because it is preferred when the number of ties is large (Nie et al., 1970). For the 21 subjects there were 274 total data points in the self-recording conditions; out of this total, 145 were ties. Therefore, the Kendall correlation was chosen.

The data for one subject in the verbal feedback group during the S-R Unaware condition could not be used to calculate a correlation because the independent observers recorded zero face-touches for all seven intervals the subject was observed. During the same seven intervals, the subject recorded zero face-touches six times and one face-touch once. In order to calculate a correlation, all zeroes were changed to one and the one face-touch recorded by the subject was changed to two.

The mean number of face-touches recorded by the independent observers and the mean frequency recorded by the subjects in each feedback group is plotted for each day under both self-recording conditions (Figures 2, 3, and 4). The graph of the data indicates little difference between the three experimental groups. The difference in reliability between the S-R Aware and S-R Unaware conditions is

not consistent. The non-specific feedback group appears to be more reliable in the S-R Aware condition than in the S-R Unaware condition, but the data from the verbal plus numerical feedback group seem to be the reverse of this. There appears to be no difference in reliability between S-R Aware and S-R Unaware conditions for the verbal feedback group.

The analysis of variance performed on the reliability data, which is presented in Table 6, confirms that there was no significant difference between the three treatment groups (F = 1.8118; df = 2, 18; $p \geq .10$). No difference in reliability of recording was found between the S-R Aware and S-R Unaware conditions (F = .0016; df = 1, 18; $p \geq .25$).

Reliability for Independent Observers

A Kendall correlation coefficient was calculated for the reliability between independent observers for each of the four conditions and for the study overall. The data used were the frequency counts of face-touching for 5-minute intervals. During the Base condition, the reliability coefficient was .8848 (N = 63). The reliability coefficient was .8380 during the S-R Aware condition (N = 30), .8913 during S-R Unaware condition (N = 31), and .8911 during the RT-Base condition (N = 30). For the study overall, the reliability of the independent observers was .9217 (N = 154).

Check on Manipulation

Upon completion of the self-recording conditions, the subjects filled out a questionnaire (Appendix B). When asked how self-recording affected the frequency of their face-touching (#5), 16 subjects correctly indicated that it produced a decrease in frequency, while four subjects thought there was no change. Of the 20 subjects that filled out the questionnaire, only one indicated that the accuracy of self-recording was affected by a lack of understanding of the rules for self-recording (#8). The subjects were also asked to rate how accurate their self-recordings were (#4) on a 5-point scale from one (not accurate) to five (very accurate). The mean rating for each of the three groups (N-S F group = 3.86; VF group = 4.29; and V+NF group = 3.83) indicates the subjects thought their recordings were accurate (4) or somewhat accurate (3). A one-way analysis of variance found no significant difference between the ratings of the groups ($F = 1.05$; $df = 2, 17$; $p \geq .25$).

Since the purpose of the study was to determine if feedback could increase the reliability of self-recording, the subjects were asked question #10, "In relation to changing your self-recording, how instrumental was the daily feedback you received in envelopes?" The subjects rated the importance of feedback on a 5-point scale with

one indicating the feedback had no effect and five indicating the feedback changed self-recording very much. The feedback was rated as being of little value in changing self-recording as is indicated by the mean ratings of the groups (N-S F group = 1.86; VF group = 2.43; and V+NF group = 2.83). A one-way analysis of variance found no significant difference among the group ratings ($F = .98$; $df = 2, 17$; $p > .25$).

As a check on the unobtrusiveness of the independent observers behind the one-way mirror, question #11 asked, "What aspects of this study, if any, aroused your suspicion?" Two subjects stated they thought an independent observer was behind the one-way mirror; two other subjects indicated that the mirror made them suspicious, but they did not specify that they thought observers were behind it. Seven subjects indicated that they were suspicious of the feedback and thought that it may be false feedback.

CHAPTER IV

DISCUSSION

Reactivity

It was found that self-recording resulted in a decrease in the frequency of the subjects' face-touching. The frequency of face-touching declined from a mean of 13.18 during baseline to 2.33 (averaged mean for both conditions) for the S-R Aware and S-R Unaware conditions. These results are in agreement with previous studies which found that self-recording results in behavior change. Self-recording has resulted in a decrease in the frequency of facial tics, out of seat, inappropriate hand raising in class, repetitive scratching, fingernail biting (Maletzky, 1974), talking out in class (Broden et al., 1971), reported hallucinations (Rutner & Bugle, 1969), and excessive eating (Mahoney, 1974). Self-recording has increased study behavior (Johnson & White, 1971; Broden et al., 1971), oral class participation (Cottman & McFall, 1972), and resisting urges to smoke (McFall, 1970; McFall & Hammen, 1971). The decrease in the frequency of face-touching is in agreement with previous studies with face-touching as the target behavior (Lipinski & Nelson, 1974; Nelson et al., in press).

A possible explanation for the change in behavior is the feedback loop hypothesized by Kanfer (1970). The feedback loop is supposedly operating when the subject

considers change in behavior to be important. When the subject self-records, he is attending to the behavior. This feedback from self-observation is compared with the desired performance criteria; if the behavior observed is at the criterion level, the subject engages in self-reinforcement. Thus, the behavior change took place because the high positive valence of the behavior resulted in self-adjustive behaviors. The results of the post-experimental questionnaire indicated the subjects were aware of the behavior change. Of the 20 subjects that filled out the questionnaire, 16 correctly indicated that self-recording resulted in a decrease in the frequency of face-touching and four subjects reported no change in frequency.

Reliability

The reliability of the subjects' self-recording was calculated in two ways. For feedback purposes, a ratio of the frequency counts of the self-recorder and the independent observer was used, i.e., the smaller frequency count was divided by the larger frequency count. This method was very sensitive to small changes in self-recording because the target behavior was of low frequency. For example, if the self-recorder recorded two face-touches and the independent observer recorded two face-touches, the reliability

was 100%. However, if during the next observed interval the subject recorded two face-touches and the independent observer recorded three, reliability decreased to 67%.

In order to perform an analysis of variance on the reliability data of the three treatment groups, reliability was also determined by calculating a Kendall correlation on the data for each subject for both the S-R Aware and S-R Unaware conditions. In contrast to the previous method of calculating reliability, the Kendall correlation is insensitive to small changes in reliability.

This indication that reliability is influenced by the method of calculation has recently been verified (Repp, Deitz, Boles, Deitz, & Repp, 1974). In their study, reliability was calculated on the same data using three different methods. Reliability was calculated using the whole session, exact agreement, and category methods. In the whole session method, interobserver reliability is calculated by dividing the smaller frequency count by the larger frequency count. When the exact agreement method is used, an interval of agreement occurs when both observers record the same frequency count. The percent agreement is calculated by dividing the number of intervals of agreement by the total number of intervals. A third method of calculating reliability is the category method. The observers are considered to be in agreement when they both record

either no responses or at least one response. It was found that the category method resulted in the highest reliability, with the whole session method producing the next highest reliability, and the exact agreement method resulting in the lowest reliability.

The results of the reliability portion of the study were not as predicted. It was predicted that the self-observers would be more reliable during the S-R Aware condition than during the S-R Unaware condition. The analysis of variance found no difference between the two conditions. It was also predicted that the V+NF group would be more reliable than the VF group which in turn would be more reliable than the N-S F group. Again, the analysis of variance found no significant difference among the three feedback groups.

The result of comparable reliability during the S-R Aware and S-R Unaware conditions is at odds with previous research utilizing self-observers and independent observers. Lipinski and Nelson (1974) found the reliability between self-observers and independent observers to be .86 when they were aware that reliability was being assessed, but this dropped to .52 when they were unaware that reliability was being assessed. In a similar study, Nelson *et al.* (in press) found reliability to be .810 when the self-recorders were aware of reliability checks and .554 when they were unaware of reliability assessment.

This drop in reliability when observers are unaware of reliability checks is comparable when reliability is calculated between independent observers. Reid (1970) found the median reliability of independent observers to be .75 when they were aware of reliability assessment but the median reliability decreased to .51 when the observers were told reliability would not be assessed. Taplin and Reid (1973) found that the mean reliability of independent observers dropped from .80 during overt reliability checks to a mean of .65 during covert reliability assessment. Romanczyk et al. (1973) also found that the reliability of independent observers was higher when they were aware of reliability checks than when they were unaware.

The failure to find a difference between the S-R Aware and S-R Unaware conditions may be due to differences between this study and other self-recording studies. In contrast to the subjects used by Lipinski and Nelson (1974) and Nelson et al. (in press), the subjects in the present study were much older and attended class after work. They may have reacted to the independent observer's coming and going much less than undergraduates. Also, unlike the Nelson et al. (in press) study, the class instructor was not connected with this study so the subjects may have viewed the study as being unimportant and the independent observer as an outsider. In addition, the class in this

study was held in the late afternoon whereas the classes used by Lipinski and Nelson (1974) were held in the morning and the class used by Nelson et al. (in press) was held at midday. Since the class was held from 4:00 - 5:30 p.m. and many of the subjects attended class after work they may have been so tired that it made no difference to them if they were being observed or not. The failure to find a difference between the S-R Aware and S-R Unaware conditions in this study may be due to these differences from previous studies acting individually or in combination. Also, a unique combination of these differences and factors common to other self-recording studies may have prevented the predicted difference between the S-R Aware and S-R Unaware conditions.

Along with the failure to find a difference between the S-R Aware and S-R Unaware conditions, there was also no difference found in the reliability of the treatment groups. It had been assumed that the value of the feedback for changing behavior was in the amount of information in the feedback. Since the V+NF group received more informative feedback on the reliability of their recording, it was predicted that this group would be the most reliable. The VF group received some information on the reliability of their recording and was predicted to be the next most reliable group. The N-S F group was predicted to be the least

reliable because their feedback provided no information as to the reliability of their recording.

One possible explanation for the failure to find a difference between the groups was that the feedback was delayed. The subjects were given feedback on their performance at the next class they attended. The course was scheduled on Tuesdays and Thursdays, so feedback was delayed at least two days.

A second reason for the lack of a feedback effect was the inappropriate pairing of verbal feedback with changes in reliability. Since previous studies had found reliability to be approximately .80 in the S-R Aware condition and .50 in the S-R Unaware condition, the changes in reliability were assumed to be small. A total of 12 verbal feedback statements were used: five were positive statements for increases in reliability, five were negative for decreases in reliability, one statement indicated no change, and one indicated high reliability (see Table 1). Since changes in reliability were assumed to be small, each of the five positive statements was paired with a 3% increase in reliability and each of the five negative statements was paired with a 3% decrease in reliability. As a result of this pairing, an increase of 13% resulted in the subject receiving the most positive feedback, except for the maintenance statement; a decrease in reliability of 13% resulted

in the subject receiving the most negative feedback. Since the subjects' face-touching behaviors were of such low frequency ($\bar{X}$ = 2.06) during the S-R Aware condition, which was the condition in which the subjects' reliability was calculated for feedback purposes, disagreement between independent observers and self-observers on only one face-touch could produce great changes in reliability. For example, if the subject recorded three face-touches during an interval and the independent observer also recorded three face-touches, reliability would be 100% and the subject would receive the maintenance statement, which was the most positive feedback and was given only for reliability of .90 or more. If during the next observed interval, the subject recorded two face-touches while the independent observer recorded three, reliability would be 67% and the subject would receive the most negative feedback statement because the drop in reliability was 33%. As a result of the inappropriate pairing of feedback statements with small changes in reliability, and the low frequency of the target behavior, the subjects often received feedback indicating substantial increases or decreases in reliability.

This great fluctuation in feedback caused many of the subjects to be suspicious of the feedback. On the post-experimental questionnaire, seven of the twenty subjects indicated that they were suspicious of the feedback

procedure. This skeptical attitude toward the feedback procedure was indicated in the subjects' response to item 10 on the questionnaire which asked, "In relation to changing your self-recording, how instrumental was the daily feedback you received in envelopes?" The subjects rated the value of the feedback on a one to five scale with a rating of one indicating the feedback was of no help and five indicating it was very instrumental in changing self-recording. The mean ratings of the groups (N-S F group = 1.86; VF group = 2.43; and V+NF group = 2.83) indicates the feedback was of little influence. A one-way analysis of variance found no significant difference between the group ratings. Thus, it appears that there was no feedback treatment effect because all three groups considered the feedback to be of little value.

These explanations for no differences in reliability among the three treatment groups were based on the assumption that informative feedback facilitates learning and performance. It may be necessary to assume that the feedback must be useful in addition to being informative. For example, in concept learning studies, subjects receiving Nothing-Wrong feedback were superior to subjects receiving Right-Nothing feedback. This difference appears to be due to the utility of the feedback. The subjects receiving Right-Nothing feedback are informed that the particular

response is correct but they do not know if their hypothesis is correct. The subjects in the Nothing-Wrong feedback group are informed that their response is incorrect and that their hypothesis is also incorrect and must be changed.

The failure to find differences among the treatment groups may have been due to the failure to provide useful feedback. The subjects in the VF and V+NF groups were informed that their reliability had increased or decreased, but not what behaviors they should engage in in order to increase their reliability or what behaviors they had engaged in which decreased their reliability.

The purpose of this study was to determine if differential feedback could make self-recording a more reliable means of data collection. In a clinical setting, a therapist may ask a client to self-record the frequency of the target behavior. However, since self-recording is unreliable the therapist may have reliability checks made by friends, members of the family, co-workers, or mechanical devices (e.g., a scale). The therapist could then use the reliability data to give the client feedback on the reliability of his self-recording, perhaps reinforcing improvements in reliability. Unfortunately, due to the problems inherent in the present study, a clinical recommendation cannot be drawn.

The design of the present study does not permit the drawing of a conclusion regarding the effect of any type of verbal feedback on the reliability of self-recorders. A self-recording baseline in which no verbal feedback was given would have been necessary in order to ascertain the effects of verbal feedback on self-recorders' reliability. The conclusion can be made, however, that varying levels of information contained in the verbal feedback are irrelevant for modifying self-recorders' reliability, at least in the conditions provided in the present study.

BIBLIOGRAPHY

Arsenian, J. Young children in an insecure situation. Journal of Abnormal and Social Psychology, 1943, 38, 225-249.

Azrin, N. H. and Powell, J. Behavioral engineering: The reduction of smoking behavior by a conditioning apparatus and procedure. Journal of Applied Behavior Analysis, 1968, 1, 193-200.

Bechtel, R. B. The study of man: Human movement and architecture. Transaction, 1967, 4, 53-56

Bolstad, O. D. and Johnson, S. M. Self-regulation in the modification of disruptive classroom behavior. Journal of Applied Behavior Analysis, 1972, 5, 443-454.

Brackbill, Y. and O'Hara, J. The relative effectiveness of reward and punishment for discrimination learning in children. Journal of Comparative and Physiological Psychology, 1958, 51, 747-451.

Broden, M., Hall, R. V., and Mitts, B. The effects of self-recording on the classroom behavior of two eighth-grade students. Journal of Applied Behavior Analysis, 1971, 4, 191-199.

Buchwald, A. M. Extinction after acquisition under different verbal reinforcement combinations. Journal of Experimental Psychology, 1959, 57, 43-48.(a)

Buchwald, A. M. Experimental alterations in the effectiveness of verbal reinforcement combinations. Journal of Experimental Psychology, 1959, 57, 351-361.(b)

Buss, A. H., Braden, W., Orgel, A., and Buss, E. H. Acquisition and extinction with different verbal reinforcement combinations. Journal of Experimental Psychology, 1956, 52, 288-295.

Buss, A. H. and Buss, E. H. The effect of verbal reinforcement combinations on conceptual learning. Journal of Experimental Psychology, 1956, 52, 283-287.

Cairns, R. B. Informational properties of verbal and nonverbal events. Journal of Personality and Social Psychology, 1967, 5, 353-357.

Dodd, D. H. and Schultz, R. F. Computational procedures for estimating magnitude of effect for some analysis of variance designs. Psychological Bulletin, 1973, 79, 391-395.

Elovson, A. C. Effect of penalty and information on error. Proceedings of the Annual Convention of the American Psychological Association, 1971, 6, 549-550.

Fixsen, D. L., Phillips, E. L., and Wolf, M. M. Achievement Place: The reliability of self-reporting and peer-reporting and their effects on behavior. Journal of Applied Behavior Analysis, 1972, 5, 19-30.

Flowers, J. V. Behavior modification of cheating in an elementary school student: A brief note. Behavior Therapy, 1972, 3, 311-312.

Goldiamond, I. Self-control procedures in personal behavior problems. Psychological Reports, 1965, 17, 851-868.

Gottman, J. M. and McFall, R. M. Self-monitoring effects in a program for potential high school dropouts: A time-series analysis. Journal of Consulting and Clinical Psychology, 1972, 39, 273-281.

Hoge, M. A. and Stockling, R. J. A note on the relative value of punishment and reward as motives. Journal of Animal Behavior, 1912, 2, 43-50.

Johnson, S. M. and White, G. Self-observation as an agent of behavioral change. Behavior Therapy, 1971, 2, 488-497.

Kanfer, F. H. Self-monitoring: Methodological limitations and clinical applications. Journal of Consulting and Clinical Psychology, 1970, 35, 148-152.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont: Brooks/Cole Publishing Company, 1968.

Lair, C. and Smith, R. K. Utilization of verbal and non-verbal feedback among college Ss. Psychological Reports, 1970, 27, 807-810.

Lipinski, D. P. The effects of differential reinforcement on reliability and reactivity of self-recorders. Unpublished doctoral dissertation, University of North Carolina at Greensboro, 1974.

Lipinski, D. and Nelson, R. The reactivity and unreliability of self-recording. Journal of Consulting and Clinical Psychology, 1974, 42, 118-123.

Lovett, H. T. The effects of various degrees of knowledge of instructional objectives and two levels of feedback from formative evaluation on student achievement. Dissertation Abstracts International, 1972, 32(9-A), 5041.

Mahoney, M. F. Self-reward and self-monitoring techniques for weight control. Behavior Therapy 1974, 5, 48-57.

Maletzky, B. M. Behavior recording as treatment: A brief note. Behavior Therapy, 1974, 5, 107-111.

McFall, R. M. Effects of self-monitoring on normal smoking behavior. Journal of Consulting and Clinical Psychology, 1970, 35, 135-142.

McFall, R. M. and Hammen, C. L. Motivation, structure and self-monitoring: Role of nonspecific factors in smoking reduction. Journal of Consulting and Clinical Psychology, 1971, 37, 80-86.

Meyer, W. J. and Offenbach, S. I. Effectiveness of reward and punishment as a function of task complexity. Journal of Comparative and Physiological Psychology, 1962, 55, 532-534.

Mueller, D. J. and Gumina, J. M. Methodological studies involving reward and punishment in human learning: 1. Effects of design and instructions. Perceptual and Motor Skills, 1972, 34, 63-74.

Nelson, R. O., Lipinski, D. P., and Black, J. L. The effect of expectancy on the reactivity of self-recording. Behavior Therapy, in press.

Nelson, R. O., Lipinski, D. P., and Black, J. L. Adult retardates: Self-recording of behaviors with different desirabilities. Paper presented at the Eighth Annual Convention of the Association for the Advancement of Behavior Therapy, Chicago, November, 1974.

Nelson, C. M. and McReynolds, W. T. Self-recording and control of behavior: A reply to Simkins. Behavior Therapy, 1971, 2, 594-597.

Nie, N., Bent, D. H., and Hull, C. H. Statistical package for the social sciences. New York: McGraw-Hill Book Company, 1970.

Patterson, G. R. and Harris, A. Some methodological considerations for observation procedures. Paper presented at the American Psychological Association Annual Convention, San Francisco, 1968.

Penney, R. K. and Lupton, A. A. Children's discrimination learning as a function of reward and punishment. Journal of Comparative and Physiological Psychology, 1961, 54, 449-451.

Polansky, N., Freeman, W., Horowitz, M., Irwin, L., Papanis, H., Rappaport, D., and Whaley, F. Problems of interpersonal relations in research on groups. Human Relations, 1949, 2, 281-291.

Reid, J. B. Reliability assessment of observation data: A possible methodological problem. Child Development, 1970, 41, 1143-1150.

Repp, A. C., Deitz, D. D., Boles, S. M., Deitz, S. M., and Repp, C. F. Differences among common methods for calculating interobserver agreement in applied behavioral studies. Paper presented at the Southeastern Psychological Association Annual Convention, Hollywood, 1974.

Romanczyk, R. G., Kent, R. N., Diament, C., and O'Leary, K. D. Measuring the reliability of observational data: A reactive process. Journal of Applied Behavior Analysis, 1973, 6, 175-184.

Rutner, I. T. and Bugle, C. An experimental procedure for the modification of psychotic behavior. Journal of Consulting and Clinical Psychology, 1969, 33, 651-653.

Siegel, L. S. and Downey, J. L. The relative effectiveness of positive and negative information feedback in a concept attainment task. Psychonomic Science, 1970, 19, 105-106.

Simkins, L. The reliability of self-recorded behaviors. Behavior Therapy, 1971, 2, 124-132.(a)

Simkins, L. A rejoinder to Nelson and McReynolds on the self-recording of behavior. Behavior Therapy, 1971, 2, 598-601.(b)

Spear, P. S. Motivational effects of praise and criticism on children's learning. Developmental Psychology, 1970, 3, 124-132.

Stuart, R. B. Behavioral control of overeating. Behavior Research and Therapy, 1967, 57, 357-365.

Talkington, L. W., Altman, R., and Grinnell, T. K. Effects of positive and negative feedback on the motor performance of mongoloids. Perceptual and Motor Skills, 1971, 33, 1075-1078.

Taplin, P. S. and Reid, J. B. Effects of instructional set and experimenter influence on observer reliability. Child Development, 1973, 44, 547-554.

Terrell, C. and Kennedy, W. A. Discrimination learning and transposition in children as a function of the nature of the reward. Journal of Experimental Psychology, 1957, 53, 257-260.

Walker, H. M. and Buckley, N. K. Effects of reinforcement, punishment, and feedback upon academic response rate. Psychology in the Schools, 1972, 9, 186-193.

Warden, C. J. and Aylesworth, M. The relative value of reward and punishment in the formation of a visual discrimination habit in the white rat. Journal of Comparative and Physiological Psychology, 1927, 7, 117-127.

Watson, D. and Tharp, R. Self-directed behavior: Self-modification for personal adjustment. Monterey: Brooks/Cole Publishing Company, 1972.

Willcutt, H. C. and Kennedy, W. A. Relation of intelligence to effectiveness of praise and reproof as reinforcers for fourth-graders. Perceptual and Motor Skills, 1963, 17, 695-697.

APPENDIX A

Instructions for Students in the Self-Recording Study

1. Each day, please fill out the data sheets completely (name, date, and number of face-touches). If you forget to record an interval, please write "forgot." Zero means no face-touches.
2. For each class day throughout the study you will be informed as to how you are doing in your self-recording. This information will be provided in an envelope attached to the data sheet. This information is on your performance so please do not discuss it with other members of the class. It is important for this research that other people's performance and feedback do not influence your recording, so please neither ask to see someone else's feedback nor show yours. Your cooperation is appreciated.

APPENDIX B

Post-Treatment Questionnaire

Name _____
Campus or
Local Address _____

1. What do you think was the purpose of this study?

2. Did you enjoy participating in this study?

1	2	3	4	5
not at all				very much

3. How much did participating in this study interfere with your class?

1	2	3	4	5
not at all				very much

4. How accurate do you think your self-recordings were?

1	2	3	4	5
not accurate	somewhat inaccurate	somewhat accurate	accurate	very accurate

5. Now that your self-recording is completed, how do you think that self-recording affected the frequency of your face-touching?

_____ no change in frequency

_____ increase in frequency

_____ decrease in frequency

APPENDIX B (Continued)

6. How certain are you that this happened?

1	2	3	4	5
very uncertain	uncertain	somewhat certain	certain	very certain

7. Describe any difficulties which you had in self-recording?

8. Do you think that the accurary of your recording was affected by your lack of understanding of the rules for self-recording listed on the bottom of the data sheets?

Yes _____ No _____

9. You were receiving daily feedback in envelopes. Briefly, what did your feedback say?

10. In relation to changing your self-recording, how instrumental was the daily feedback you received in envelopes?

1	2	3	4	5
not at all				very much

11. What aspects of this study, if any aroused your suspicion?

12. Comments, complaints, or suggestions, for improvements in this type of study:

APPENDIX C

Feedback Forms

Form 1

Good work! Your data sheet for last class was filled out completely. Thank you for your cooperation.

This is the form the subjects in the non-specific feedback group received for turning in complete data sheets.

Form 2

Date _____

During the last class,

there _____

in your reliability.

This is the form subjects in the verbal feedback group were given. The words written in the blank were determined by the change in the subject's reliability.

APPENDIX C (Continued)

Form 3

Date _____

Your reliability for the last class was _____

This is an increase (decrease) of _____ from the class meeting prior to the last class.

Subjects in the verbal plus numerical feedback group received feedback on this form. The current reliability and change in reliability were placed in the blanks provided The corresponding verbal statement was also added below the printing.

Form 4

There were _____ 5-minute intervals left blank on your data sheet for the last class. Each day, please fill out the data sheets completely (name, date, and number of face-touches). If you forget to record an interval, please write "forgot." Zero means no face-touches.

Any subject who turned in an incomplete data sheet was given this form in the next class.

APPENDIX C (Continued)

Form 5

Since you were absent from the last class meeting it is important that you fill in all intervals on the data sheet today so that the study can be completed before the semester ends. Thank you for your cooperation.

Any subject who was absent received this form on their return to class.

APPENDIX D

Tables and Figures

Table 1

Verbal Feedback Given VF and V+NF Groups

% of Change in Reliability	Verbal Statement
+ 1-3	There has been a small improvement in your reliability.
+ 4-6	There has been an improvement in your reliability.
+ 7-9	There has been a good improvement in your reliability.
+ 10-12	There has been a great deal of improvement in your reliability.
+ 13 or more	There has been an outstanding improvement in your reliability.
No change	There has been no change in your reliability.
- 1-3	There has been a slight decrease in your reliability.
- 4-6	There has been a decrease in your reliability.
- 7-9	There has been a substantial decrease in your reliability.
- 10-12	There has been a very great decrease in your reliability.
- 13 or more	Your reliability for last class was very poor.
Maintenance statement	Outstanding! The reliability of your recording is exceptional.

Table 2

Mean Face-Touching Frequency of Subjects

Subjects		Conditions: Base	S-R Aware	S-R Unaware	RT-Base
Non-Specific Feedback Group	S_1	23.43	2.57	4.00	23.14
	S_2	16.00	.86	2.71	8.14
	S_3	13.85	2.57	3.00	8.86
	S_4	10.85	1.00	1.43	6.29
	S_5	10.43	3.71	2.43	8.14
	S_6	9.00	3.00	4.29	7.86
	S_7	8.57	1.29	3.43	7.00
		$\bar{X}$ = 13.16	$\bar{X}$ = 2.14	$\bar{X}$ = 3.04	$\bar{X}$ = 9.92
Verbal Feedback Group	S_8	23.29	3.71	3.71	13.14
	S_9	18.29	1.57	1.43	7.86
	S_{10}	12.28	2.86	2.43	8.14
	S_{11}	11.00	1.57	.86	9.86
	S_{12}	9.85	.29	.14	6.00
	S_{13}	9.00	3.43	3.57	10.43
	S_{14}	8.43	1.00	2.43	5.71
		$\bar{X}$ = 13.16	$\bar{X}$ = 2.06	$\bar{X}$ = 2.08	$\bar{X}$ = 8.73
Verbal + Numerical Feedback Group	S_{15}	22.14	2.71	6.00	16.29
	S_{16}	19.14	2.43	3.00	6.43
	S_{17}	12.29	2.57	1.86	10.00
	S_{18}	11.28	3.29	2.29	10.57
	S_{19}	9.86	.43	3.43	8.43
	S_{20}	9.43	1.71	1.43	8.57
	S_{21}	8.29	.71	.86	13.29
		$\bar{X}$ = 13.20	$\bar{X}$ = 1.98	$\bar{X}$ = 2.70	$\bar{X}$ = 10.51
		$\bar{X}$ = 13.18	$\bar{X}$ = 2.06	$\bar{X}$ = 2.61	$\bar{X}$ = 9.72

Table 3

Analysis of Variance for Face-Touching Data

Source	df	SS	MS	F
Between Ss				
Treatments (A)	2	42.75847	21.37923	.1242
S(A)	18	3099.48000	172.19330	
Within Ss				
Conditions (B)	3	13114.50000	4371.50000	74.9122*
Repeated Observations (C)	6	315.62890	52.60481	1.8740
A X B	6	61.15234	10.19206	.1747
A X C	12	426.43090	35.53590	1.2659
B X C	18	288.38280	16.02126	.6134
A X B X C	36	912.83080	25.35640	.9709
SB(A)	54	3151.16800	58.35497	
SC(A)	108	3031.65900	28.07091	
SBC(A)	324	8461.80900	26.11668	
Total	587	32905.80021		

*$p < .01$

Table 4

Newman-Keuls Test for B (Experimental Conditions)

	S-R Aware 2.06	S-R Unaware 2.61	RT-Base 9.72	Base 13.18	r	Critical Value
S-R Aware	2.06	.55	7.66*	11.12*	4	7.72
S-R Unaware	2.61		7.11*	10.57*	3	7.13
RT-Base	9.72			3.46	2	6.27

*$p < .01$

Table 5

Reliability Between Subjects and Independent Observers (Kendall Correlation)

		S-R Aware	S-R Unaware
Non-Specific Feedback Group	S_1	.3158	.4326
	S_2	.4537	.5008
	S_3	.9701	.8903
	S_4	.8427	.8767
	S_5	.6489	.7233
	S_6	.5477	.9234
	S_7	.6860	.1667
Overall Reliability		.6801	.6690
Verbal Feedback Group	S_8	.7895	.8293
	S_9	.5407	.9393
	S_{10}	.7515	.9234
	S_{11}	.5294	.7882
	S_{12}	.7303	1.0000
	S_{13}	.8903	.6500
	S_{14}	.8281	.7515
Overall Reliability		.7841	.8706
Verbal + Numerical Feedback Group	S_{15}	.7790	.3338
	S_{16}	.6482	.5564
	S_{17}	.9147	.7059
	S_{18}	.7233	.4104
	S_{19}	.4201	.6299
	S_{20}	.9075	.7432
	S_{21}	.6211	.8058
Overall Reliability		.7477	.6098

Table 6

Analysis of Variance for Reliability Data

Source	df	SS	MS	F
Between Ss				
Treatments (A)	2	.16413580000	.08206791000	1.8118
S(A)	18	.81534780000	.04529710000	
Within Ss				
Conditions (B)	1	.00004821476	.00004821476	.0016
A X B	2	.09699517000	.04849759000	1.6059
SB(A)	18	.54359540000	.03019974000	
Total	41	1.62012200000		

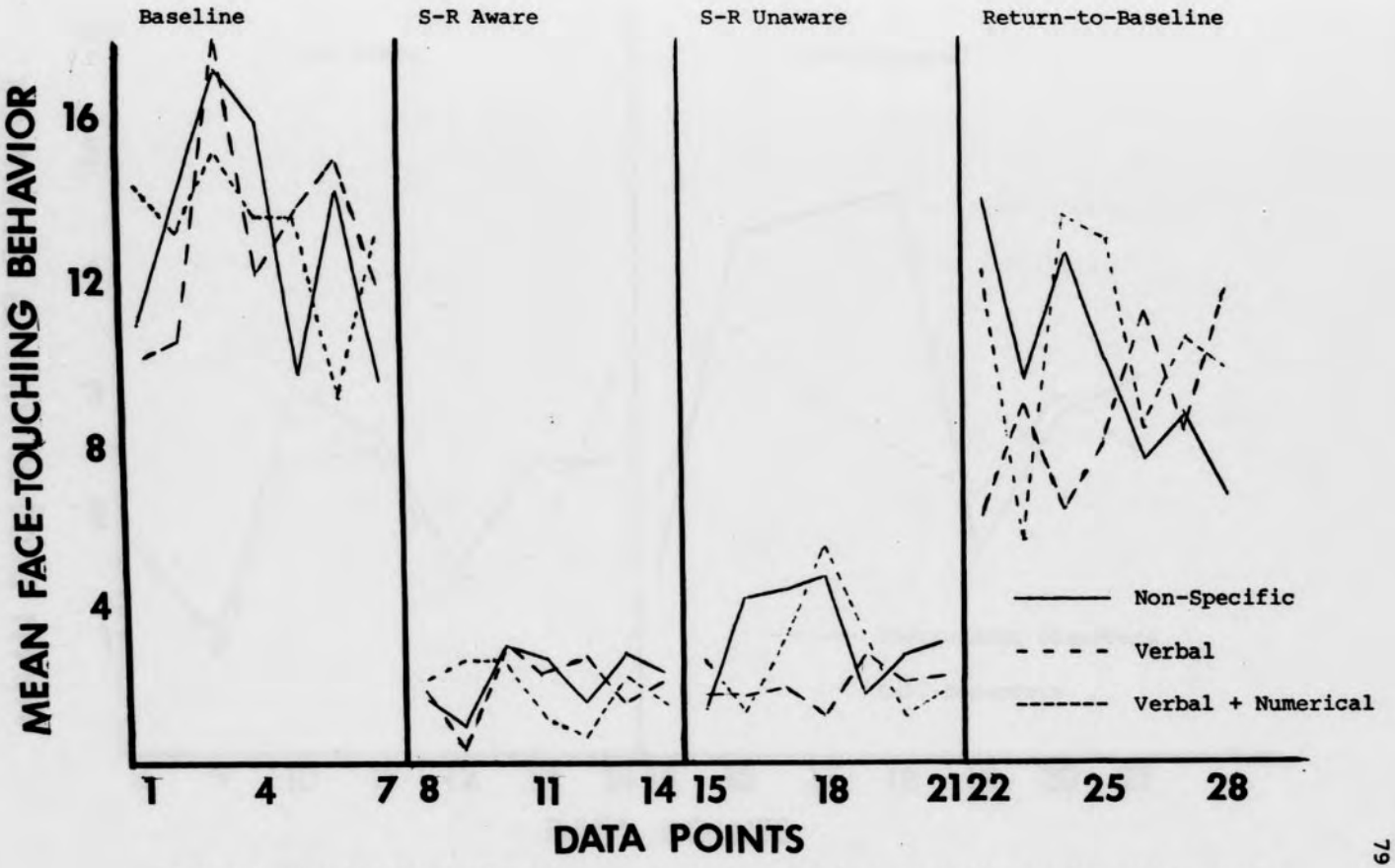

Figure 1. Mean face-touching behavior for the three groups of subjects per 5-minute interval. The S-R Aware and S-R Unaware conditions were interspersed among each other but are presented separately.

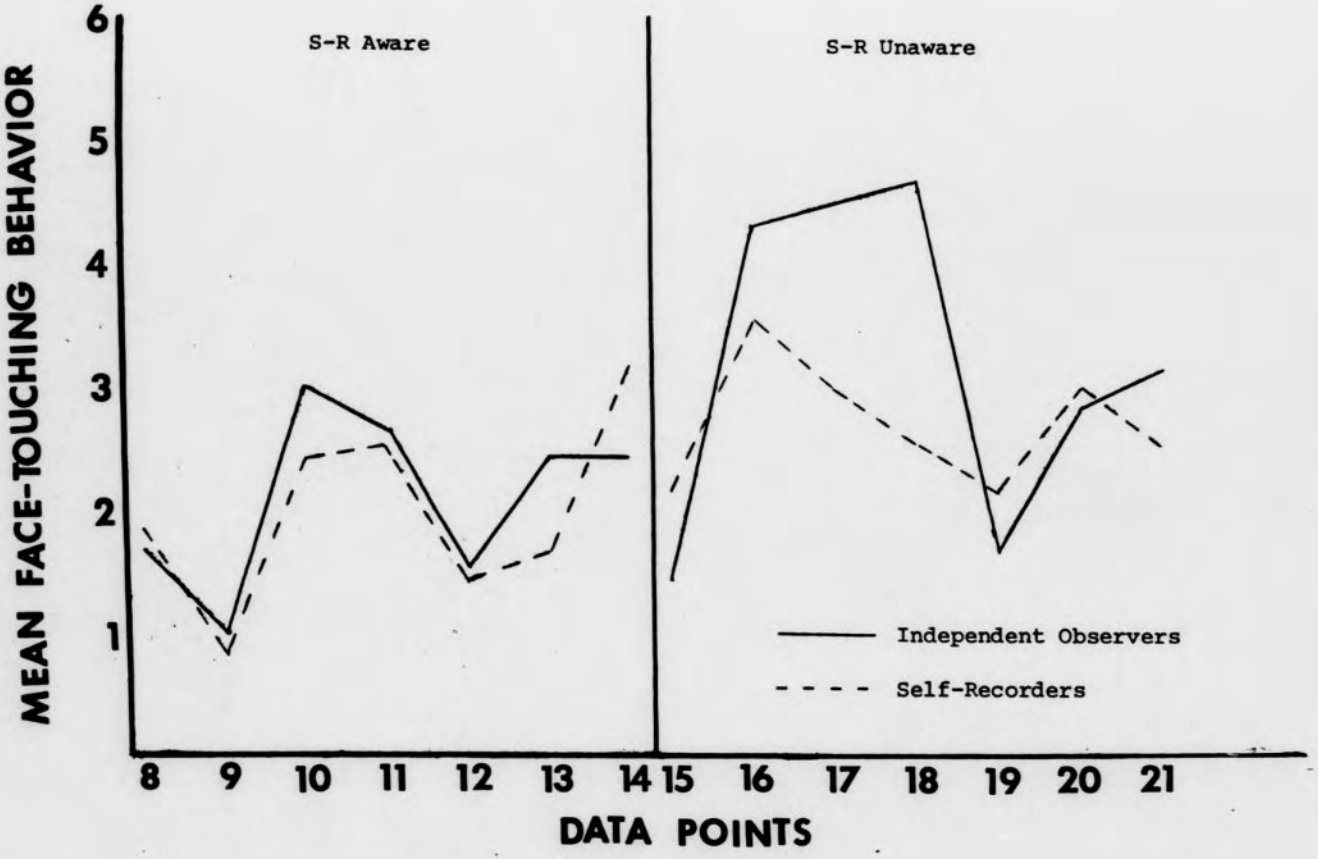

Figure 2. Mean face-touching behavior recorded by the independent observers and self-recorders in the non-specific feedback group per 5-minute interval. The S-R Aware and S-R Unaware conditions were interspersed among each other but are presented separately.

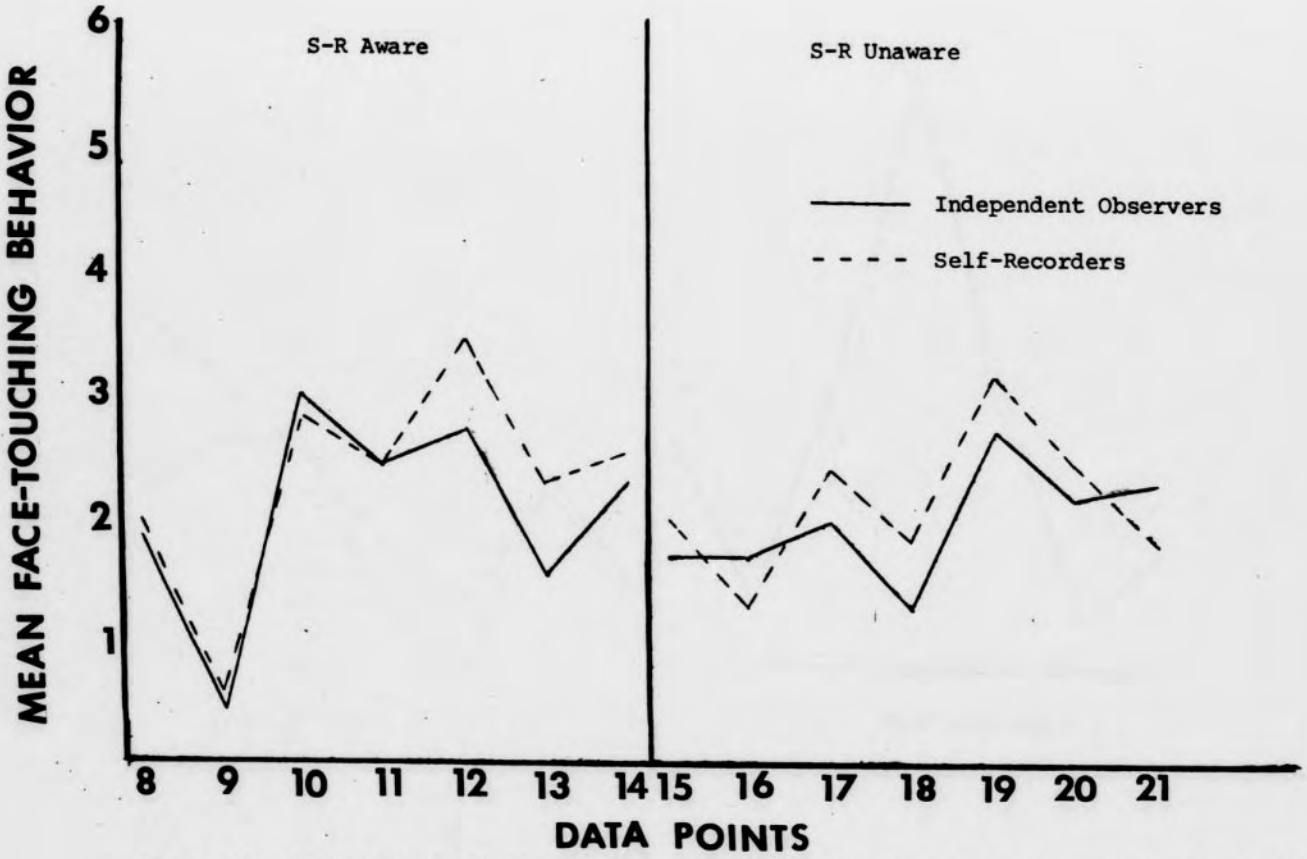

Figure 3. Mean face-touching behavior recorded by the independent observers and self-recorders in the verbal feedback group per 5-minute interval. The S-R Aware and S-R Unaware conditions were interspersed among each other but are presented separately.

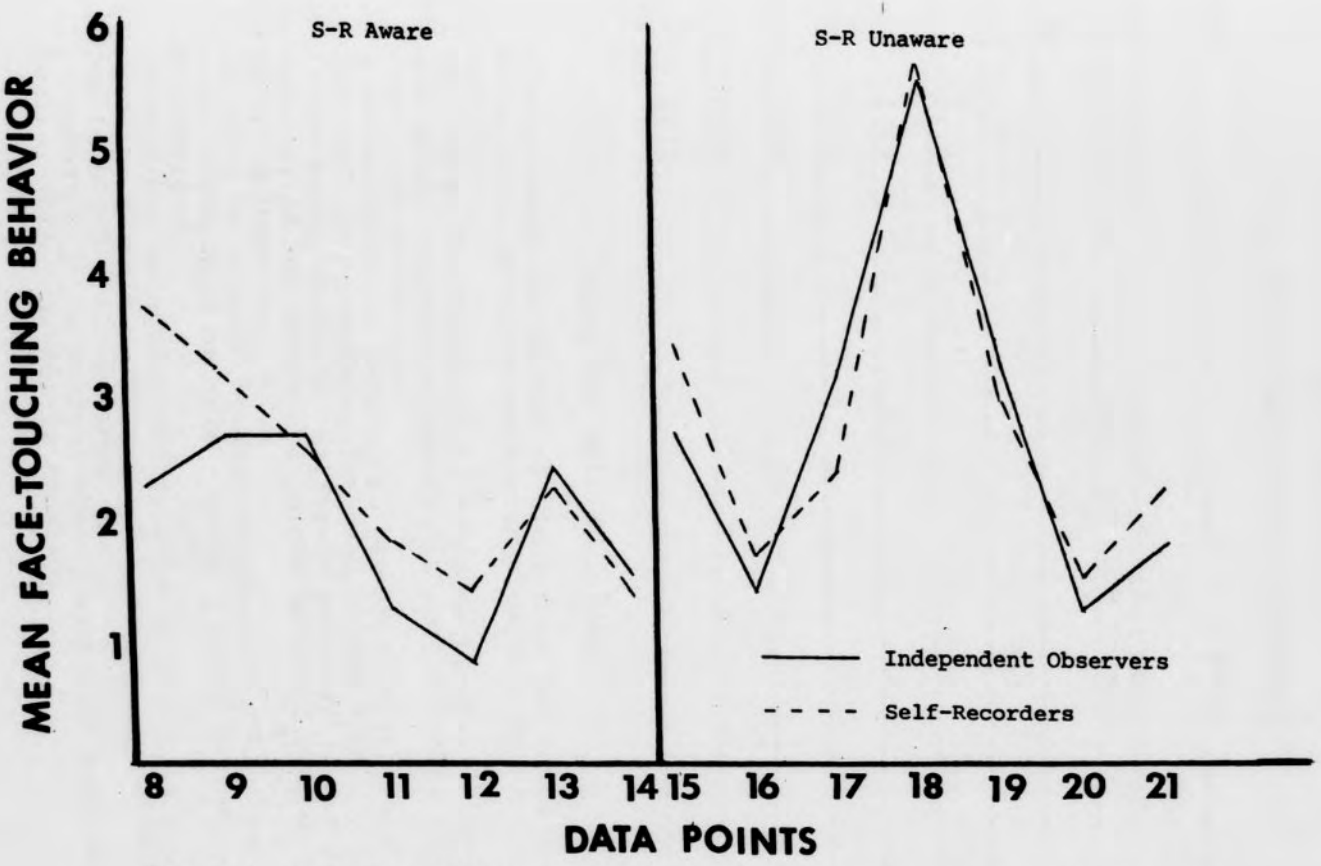

Figure 4. Mean face-touching behavior recorded by the independent observers and self-recorders in the verbal plus numerical feedback group per 5-minute interval. The S-R Aware and S-R Unaware conditions were interspersed among each other but are presented separately.

APPENDIX E

Data Sheets Used by Self-Recorders

Name _____ Date _____

Time Interval (5-minute blocks	Number of times face, head, hair, or neck touched
4:00 - 4:04	
4:05 - 4:09	
4:10 - 4:14	
4:15 - 4:19	
4:20 - 4:24	
4:25 - 4:29	
4:30 - 4:34	
4:35 - 4:39	

Rules for Self-Recording

1. Simply count the number of times that you touch your face, head, hair, or neck with your hand or pen (pencil).
2. This sheet is divided into 5-minute blocks. Please make every attempt to coordinate your counting with these time intervals.
3. A new behavior is defined as breaking contact with the face and then returning the hand to the face.
4. If two hands touch two different parts of face = 2 behaviors; if two hands are joined to touch one part of face = 1 behavior.
5. If there is a pause between touches = 2 behaviors: e.g., right hand on nose - pause - right hand on nose = 2 behaviors.
6. Objects may be equivalent to face or hands: e.g., touching glasses with hand = 1 behavior; or touching pen to mouth = 1 behavior.
7. If behavior is in process when observation interval begins, this behavior is counted.

APPENDIX F

Data Sheet Used by Independent Observers

Name of Observer _____

Name of Reliability Checker _____

Date _____

Study Condition: Pretest; SR-K; SR-U; Post-test
5-minute intervals:

S	1	2	3	4	5	6	7	8	9	10	11	total