THE EFFECTS OF DIFFERENTIAL REINFORCEMENT ON RELIABILITY AND REACTIVITY OF SELF-RECORDERS

by

David Paul Lipinski

A Dissertation 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
Doctor of Philosophy

Greensboro
1974

Approved by

Rosemary O. Nelson
Dissertation Adviser

APPROVAL PAGE

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

Dissertation Adviser Rosemary O. Nelson

Committee Members P. Scott Lawrence

Robert G. Eaves

Jacqueline W. Mackelcan

Robert G. Jones

March 1, 1974

Date of Acceptance by Committee

LIPINSKI, DAVID PAUL. The Effects of Differential Reinforcement on Reliability and Reactivity of Self-Recorders. (1974) Directed by: Dr. Rosemery O. Nelson. Pp. 79.

The purpose of the present study was to investigate the possibility of enhancing the effectiveness of self-recording as both an assessment device and as a therapeutic tool. Three hypotheses were investigated. First, differentially reinforcing subjects for increments in reliability with the data recorded by independent observers would increase the subjects' reliability. Second, differentially reinforcing subjects for decrements in the behavior that was being self-recorded would result in decreases in that behavior. Third, reliability would be lower when self-recorders were unaware that reliability was being assessed than when they were aware of the assessment.

The experimental design was a 2x7x9 factorial design with subjects nested in the two treatment groups and repeated across the seven experimental conditions with nine observations under each condition. Twenty college students in classroom settings were differentially reinforced for either increases in reliability of self-recorded data or decreases in the behavior that was being self-recorded. The seven experimental conditions were: baseline, baseline aware, baseline unaware, self-recorder aware, self-recorder unaware, and return-to-baseline I and II.

The results supported the view that self-recording can serve two separate functions: that of a method of data collection, and that of a therapeutic tool. These data indicated that it was possible to differentially reinforce the therapeutic function over the assessment function, and conversely. That is, subjects who were reinforced for increasing their reliability increased their reliability without reducing their behavior significantly more than the subjects reinforced for reducing their target behavior. However, the subjects reinforced for a decrement in their target behavior reduced their target behavior, while remaining unreliable, more than the self-recorders reinforced for increasing their reliability. The data suggested that the effects of self-recording may only be temporary. The data also supported the hypothesis that self-recorders would be more reliable when they were informed of the reliability check than when they were unaware of the assessment.

The implications of the above findings are numerous. First, self-recording can be used as a method of data collection. Since self-recording is likely, however, to be unreliable, every attempt should be made to obtain periodic independent behavior ratings. The reliability of self-recording can possibly be enhanced by having independent observers reinforce self-recorders for increments in their reliability. Second, self-observing is a reactive

technique that can be used to therapeutically increase appropriate behavior or to decrease inappropriate behavior. The reactive effects of self-observing may be heightened by reinforcing the self-recorders for further changes in the target behavior. Finally, the temporary effects of self-monitoring noted in the present study and in previous studies suggest that such variables as the strength of the target behavior, the duration of self-recording, the nature of the target behavior, and the subjects' motivation for change require further investigation.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Dr. Rosemery O. Nelson for her guidance throughout the course of this dissertation and the research related to it. Sincere thanks also go to Dr. Robert Eason, Dr. Jacquelyn Gaebelein, Dr. Robert Jones and Dr. P. Scott Lawrence for their helpful suggestions.

In addition, the author would like to thank his wife, Judith, for her assistance in typing the manuscript. Finally, the author would like to take this opportunity to thank his parents, Mr. and Mrs. Stanley Lipinski, for their help throughout the author's academic career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

One characteristic which differentiates behavior therapists from less empirically-minded clinicians is an emphasis on the systematic collection of data both for research purposes and for individual therapy cases. The utility of such data collection lies in enabling the experimenter or therapist to establish an operant level of the behavior that is to be changed, and consequently to analyze the effects of treatment. According to Goldfried and Kent (1972), the most direct and non-inferential approach to data collection is behavioral sampling of the individual's actual responses in a realistic situation within the natural environment.

Since adequate mechanical recording devices have not been developed for the complex social behaviors seen in the natural environment, researchers have developed definitions of the specific behaviors with which they are concerned and have trained observers to record these behaviors. A specific behavior is described as an event that can be observed and recorded; in addition, the experimenter specifies the topographical extensions of the behavioral category so that one can easily decide if a behavior belongs within a given category.

Given the extensive use of naturalistic observations in behavior modification, a legitimate concern pertains to methodological problems which may influence the reported results. There appear to be three main categories into which these methodological problems fall: first, procedural problems in observations; second, potential observer bias; and third, the reactive nature of "being observed." The present paper will discuss each of these categories and conclude with a discussion of self-recording, illustrating the parameters of self-recording that require further research.

Procedural Problems in Observations

Loss of Information by Using Behavioral Codes

An initial method of recording human behavior in situations was simply writing an account of what the subject was doing (e.g., Wright, 1967). This method required the observer to direct most of his attention to the actual recording of the behavior (e.g., writing) which restricted the amount of attention directed toward the subject's behavior (Mischel, 1968). An improvement over the previous method is the use of behavioral codes by O'Leary, Romanczyk, Kass, Dietz, and Santogrossi (1971) and by Patterson and Cobb (1971). In both cases, the behaviors of particular interest (target behaviors) are delineated into discrete

categories represented by coded symbols and are clearly defined; for example, samples of behaviors which are included and excluded from each class of behaviors are specifically stated. However, there are differences between the O'Leary et al., and the Patterson and Cobb codes. The O'Leary et al. code uses pre-coded data sheets which simply require the observer to mark the appropriate symbols as the behaviors occur, whereas the Patterson and Cobb code has the observer write the appropriate symbols as the behaviors occur. An advantage of the Patterson and Cobb code is its provision for systematically recording not only the subject's behaviors, but also temporal sequences of the behaviors of other persons in the situation. These temporal sequences are pre-requisite to determining the functional relationship of antecedent and consequent stimuli to the target behavior. At the present time, the O'Leary et al. code does not provide for systematic recording of behavioral interactions.

Nonetheless, there remain certain problems associated with these approaches. First, any analysis of the "behavioral stream" (Wright, 1967) requires an a priori selection of behaviors which will be recorded. Other behaviors which occur and which do not fall within this selection are lost from analysis. Second, any particular class of behavior is usually recorded only once during each interval. For

example, if 20-second intervals are being used, and if a disruptive behavior occurred after the first 5 seconds, then the symbol appropriate for that behavior could not be circled again for the remaining 15 seconds. In other words, the duration, number of occurrences within that interval, as well as the specific time of onset of each type of behavior cannot be recorded. More precise information would be necessary in order to determine the nature of any systematic covariations of a subject's behavior with another person's behaviors, or with changes in the environment.

The Appropriate Time to Discontinue Recording Data

Another problem concerning the data sample is the extent to which it is representative of the population of behaviors both across time and across situations. Currently, there are no standard criteria for assuming representativeness and thus for terminating data recording. Specifically, how many data points should be included in baseline, or how low should the variance be in order to assume a stable estimate of behavior? Sidman (1960) suggested that baseline recording should continue until the baseline behavior is stable within a 5% range. As Simkins (1969) has noted, what if one does not have the requisite technology available to reduce the variability of the target behavior to 5%? There is also the possibility, to be

discussed in detail below, that an extended collection of baseline data may actually produce changes in the target behavior, due to reactive properties of observers (McNamara and MacDonough, 1972).

Patterson and Gullion (1968) suggested that consistent data are more readily achieved by recording behavior at the same time of the day, presumably in the same situation. Yet, there is no guarantee that an accurate count of behaviors has been obtained merely by limiting situations in which observations are taken. A problem related to this approach is the differential considerations required by behaviors which have high versus low base rates (Patterson & Harris, 1968). It is possible that low base rate events occur in a relatively limited set of situations, and these may occur only at widely spaced intervals. In these situations, the experimenter would be less concerned with situations and more concerned with time, specifically, the amount of observer time required to establish stable estimates. On the other hand, high base rate events (i.e., writing) may occur in many settings. In order to obtain an adequate sampling of the differences in these settings, there must be observations in multiple settings as well as over a period of time; both time and situational variables would be important in determining the adequacy of the observational sample.

The Method by Which Reliability is Calculated

A demonstration of high reliability is essential for concluding that a strong relationship exists between the behavior emitted by the subject and the behavior recorded by the observer. High reliability means that there is a high level of agreement between two or more observers who are simultaneously recording the same behavioral sequence, utilizing the same recording procedure. Recently, Johnson and Bolstad (1973) advocated the use of terms "observer accuracy" and "observer agreement" instead of the term "reliability," since "reliability" could be taken to refer to the concept of reliability as used in traditional test theory. Observer agreement is obtained by comparing the scores of two or more observers, and observer accuracy represents a comparison of an observer's score with some established criterion. The use of the terms "observer accuracy" and "observer agreement," instead of "reliability," would minimize confusion with the way reliability is used in traditional test theory. However, the term "reliability" is nonetheless frequently used by investigators in this area, and is used in the present paper to maintain consistency with the general usage in the observation literature.

Dependent upon the kind of data being recorded, there are several alternative ways of calculating reliability;

intrinsic to each method are certain assumptions and problems. For example, when using a time sampling method of observation, the level of reliability reported for various behaviors varies with the number of behavior categories being recorded. For instance, if only two behavioral categories are utilized, one would expect a higher reliability than if six categories were employed, since there would be more agreement on the specific category in which the behavior fell by chance alone. Mash and McElwee (in press) have recently provided support for this view. These authors found that observers using a four-category coding system were more accurate than observers using an eight-category system. The greater complexity of the eight-category system was offered as an explanation for the lower performance of the observers in the eight-category group.

The reliability coefficient that is obtained may also differ depending on whether or not the absence of the target behavior during an interval was recorded as an agreement, absence meaning that neither observer recorded the behavior in a given interval. If such absences are included as agreements, then the reliability coefficient would be changed. For example, suppose that there are 15 intervals with seven agreements, five disagreements, and three absences of the target behavior. When the absences are not included as agreements, the reliability coefficient

is equal to .58. On the other hand, when absences are included, the coefficient is equal to .67.

A final problem related to the issue discussed by Johnson and Bolstad (1973) reflects on the validity of recorded data, given its possibly high reliability. An assumption is made that if an agreement is recorded between two observers, then both are attending to and recording the same stimulus. Thus, if there is high reliability, it is assumed that the data recorded have high validity with the actual behaviors. However, this is not necessarily the case. It is possible for two observers to simultaneously record four instances of a behavior during a certain interval, but with the subject emitting eight instances of the target behavior. The reported reliability would be 100% for that interval but in actuality neither observer accurately recorded the target behavior.

Potential Observer Bias

An observer is biased when his report of the observational data is influenced by factors other than the occurrence of the target behaviors. Some factors which seem to influence the reports of observers are as follows: knowledge of expected results, evaluative feedback from the experimenter, and knowledge that reliability observations are being made.

Rosenthal (1963, 1966) reported several studies which indicated that the observer's perceptions and interpretations are influenced by his knowledge of expected results. Despite methodological inadequacies, Scott, Burton, and Yarrow (1967) illustrated the influence of observer expectation upon reported observations, by showing that the data from observers informed about experimental conditions showed stronger support for the experimenter's hypothesis than data from uninformed observers.

Kass and O'Leary (1970) showed that observer expectations may influence their recorded data. In a laboratory study, regardless of the fact that all observers were recording behavior from the same video tape, they found that the group told to expect a decrease in the frequency of disruptive behavior recorded a sharper decrease than either a group told to expect an increase in the frequency of disruptive behavior, or a group given no specific expectation. Kent, O'Leary, Diament, and Dietz (1973) designed a study to investigate the effects of expectation on the observational recordings of trained observers. The results demonstrated that only the observers' global subjective evaluations of the changes in the subjects' behavior were modified in the direction of their expectancies. The objective data that they recorded were not significantly influenced by these expectancies. In addition, Skindrud (1973)

either informed or did not inform observers of the normal or deviant status of families being observed. The results showed that no significant differences occurred between the informed and uninformed observers during either baseline or intervention observations of the deviant families as compared with the normal families. These latter two studies have somewhat diminished the concern that observer expectancy produces biased observations.

In addition to expectancy of results, a second source of potential observer bias is evaluative feedback from the experimenter to whom the observer reports his data. O'Leary, Kent, and Kanowitz (1972) have reported that verbal comments by the experimenter can differentially reinforce observer reports of decreases in behavior in preselected behavioral categories.

A third source of potential observer bias is knowledge that reliability observations are being undertaken. Reid (1970) presented results indicating that the level of reliability obtained depends on whether or not the observers realize reliability is being computed. Observers aware that reliability was being assessed obtained median reliabilities of .75; but, when informed reliability would not be assessed, the median reliability dropped to .51.

In a study utilizing a classroom situation, Romanczyk, Kent, Diament, and O'Leary (1973) found that reliability was

lower when observers were not informed that reliability was being computed than when the observers were instructed that a reliability assessment was being performed. These investigators also reported that knowledge of which observer is performing the assessment significantly affected the reliability level, and that the frequency of behaviors recorded was higher when observers knew that reliability was being assessed. Thus, when reliability is not being assessed, the data would be biased by underestimating the occurrence of target behaviors. Possibly the presence of the reliability checks induces the observers to pay more attention to the situation. A phenomenon which might be the result of all three sources (and others) of observer bias is one which O'Leary and colleagues have labeled observer "drift." This phenomenon refers to a random fluctuation, over time, in the observational criteria used by groups of observers. In other words, discussion of the differences in recording among the observers modifies their interpretation of the behavioral code to more closely match each other. O'Leary (1973) has suggested that when more than one group of observers is used, different modifications of the observational code may emerge and these must be differentiated from possible systematic biases due to observer expectations.

Reactive Nature of the Observation Process

An observer is reactive with a situation when his presence causes actual behavioral changes in the subject. O'Leary et al. (1971) have stated that the observer should become as neutral a stimulus as possible by minimizing the interaction between the observer and the observed. However, it appears that an observer is not a neutral stimulus. The presence of an observer disrupted the behavior of delinquent children (Polansky, Freeman, Horowitz, Irwin, Papanis, Rappaport, & Whaley, 1949) and prolonged the negative reactions of nursery school children in an unfamiliar situation (Arsenian, 1943). Bechtel's (1967) results may support the assumption that observers are an aversive stimulus. Data were collected on museum visitors who knew they were being observed and on those who did not know; people who knew they were being observed spent less time in the room and made less movement around the room than people who did not know their movement was being recorded. Similarly, White (1973) reported that introducing an observer into the room reduced the activity level of people within that room.

In their work with families, Patterson and Harris (1968) indicated that the presence of an observer constituted a stimulus producing different behavior from that obtained when no observer is present. If the observer is

judged to be an aversive stimulus, there may be an increase in escape and avoidance behaviors, such as time spent in the bathroom or at the store, or a suppression of usual behaviors. Conversely, the family members may feel obligated to interact more frequently. In either case, the presence of the observer per se reduces the validity of the observations. It then becomes an empirical question if and when habituation to the observer occurs. Johnson and Bolstad (1973) have pointed out that the reactive effects of being observed depend on many factors such as the level of conspiciousness of the observer, individual differences of the subjects, personal attributes of the observers, and the rationale that is given for the observations.

Closely related to the issue of reactivity is the view presented by Johnson and Lobitz (1972). These investigators suggested that people being observed would modify their behavior to correspond with the demand characteristics of the experimental setting (or naturalistic setting). They found that parents, in response to instructions, could control the deviancy level of their children. An implication of this is that a parent might attempt to make the child appear deviant (either by behaving in ways that actually produce deviancy or by reporting greater deviancy than actually occurs) during baseline so that

treatment will be justified, and might at conclusion of treatment, make the child appear improved in order to please the therapist and to justify termination of the treatment.

Self-Recording

The discussion of the reactive nature of observation has thus far been limited to situations where an observer is recording the subject's behavior. Recently, some behavior modification programs have required subjects to self-observe and to self-record their own behavior (e.g., Goldiamond, 1965; Stuart, 1967). Self-recording seems to be as reactive as other types of observations; the individual who is recording his own behavior is aware not only of the exact target behavior, but also of the purpose of the observations. The reactive nature of self-recording has two components: actual changes in behavior as a function of self-recording, and the reliability of self-observers.

The reactivity of self-recording has been shown in several experiments. McFall (1970) had one group of subjects record their frequency of in-class smoking behavior and another group, in-class abstinence (the act of resisting an urge to smoke). He discovered that self-recording of in-class smoking increased its frequency while

self-observations of in-class abstinence decreased the frequency of smoking. In a later study, McFall and Hammen (1971) had smokers who wished to decrease their smoking rate self-record either frequency of smoking, the number of times they resisted an urge to smoke, or the number of times they were unable to resist an urge to smoke. Despite the fact that all groups were told to record different behaviors, a reduction in smoking rate was noted, and there were no significant differences between groups.

The disparate results noted between McFall's studies may be explained by the fact that the subjects in McFall's initial study were students attending a college class instructed by the experimenter who facilitated smoking by systematically modeling smoking behavior. In his later study, McFall used subjects who attended a smoking clinic and who had previously indicated that they wanted to terminate their smoking behavior. The subjects were also given the expectancy that following the program to which they were assigned would end their smoking.

Johnson and White (1971) also affirmed that self-recording is reactive to the extent that it can change behavior in a desired direction. Students who were asked to self-record their studying behavior obtained significantly higher grades than a control group who were not

self-recording. Subjects who self-recorded dating activities also obtained higher grades than the control group but lower grades than the students who self-recorded study behavior; however, these differences were not significant. Johnson and White (1971) attempted to explain these results in the dating group by noting that self-recording of a behavior like dating helped the subjects evaluate the way that their time was being invested in various activities. Also, self-recording of one behavior may have generalized to another behavior (i.e., study), if they found self-observation to be a useful method for organizing their time. In fact, some subjects in the dating group actually reported that they self-recorded their study time. In addition, Broden, Hall, and Mitts (1971) found self-recording increased an appropriate behavior (studying) and decreased an inappropriate behavior (talking out) in a classroom setting. Rutner and Bugle (1969) used self-monitoring to reduce hallucinatory behavior, and Ernst (1973) utilized self-recording and counter-conditioning to decrease the frequency of mouth-biting. Recently, Maletzky (1974) described five individual case studies in which self-recording was used to decrease inappropriate motor behaviors (e.g., repetitive scratching, fingernail biting, facial tics, hand waving, and out-of-seat). These results support the notion that self-recording is reactive.

Even within a single behavior the effects of self-recording can be differentially reactive, dependent upon the exact instructions given, as shown by Gottman and McFall (1972). Their results showed that the reversal of the two self-monitoring instructions produced a cross-over effect. Subjects told to monitor talking increased their talking, but when instructed to monitor non-talking, they decreased their talking. Two other dependent measures showed minimal or no effects -- grades and office visits -- suggesting that the effects (reactive) of self-recording are specific to the behavior being monitored. Further evidence for the reactive effects of self-recording can be found in the study by Nelson, Lipinski, and Black (1974). The major purpose of this investigation was to manipulate the direction of the behavior change produced by self-recording by giving self-recorders different expectancies of the direction of behavior change. Regardless of the expectancies given to the self-recorders, all groups decreased their face-touching. It appeared that the manipulation of expectancies was a weak variable that failed to alter the direction of change produced by the self-recording of face touches.

Although self-recording has been found to be reactive, the direction of the changes produced by the process of self-recording differ, perhaps as a function of one or more

variables. One possible variable that might influence the direction of reactive behavior changes in self-recorders could be the differential instructions given the self-recorders (Kanfer, 1970). McFall (1970) found that subjects who were told to record the number of cigarettes smoked increased their smoking behavior, whereas subjects who were told to count the number of times they resisted smoking urges decreased smoking. However, McFall and Hammen (1971) found that all self-observers decreased their smoking despite varying sets of instructions. Gottman and McFall (1972) found that subjects who were told to record the number of times they talked during a class discussion increased their classroom participation, while subjects told to monitor non-talking decreased their talking. Thus, the specific instruction given to self-recorders controlled the direction of behavior change. Another possible reason for reactivity may be that self-recording calls attention to a behavior and if negative value judgments are attached to the behavior by others, the behavior would decrease in frequency because such a decrease would be self-reinforcing and perhaps positive value judgments would increase the behavior. Therefore, it is appropriate to reduce smoking, weight, or classroom disruptions while appropriate to increase studying or class participation. A third possible variable may be

that the materials used for self-recording may function as discriminative stimuli for producing behavioral change. If the behavior were perceived as undesirable, the self-recording materials could alert the person prior to the execution of the behavior which would interrupt the undesired behavior sequence. If the behavior was judged as desirable, the materials used for self-recording could prompt the person to engage in this behavior. In addition, if the subject's beliefs conflict with his self-recorded data, the conflicting self-perception or dissonance may result in behavior change (Kanfer, 1970). Other variables such as the client's motivation to change, the incompatibility between the observing response and the symptomatic behavior, and the reinforcing consequences of the observed behavior might influence behavior change via self-recording (Kanfer, 1970). Despite the uncertainty regarding the variables which produce the reactive effects on the self-recorded behavior, the results suggest that the reactivity of self-recording may be used to produce desired behavior change.

A problem encountered when reviewing self-recording studies is the failure to assess reliability. Simkins (1971) stated that there has been a general neglect in assessing the self-recorder's reliability and that there is a need to assess self-observer bias in the recording

of specified variables. Only one study (McFall, 1970) reported a correlation (.61) between the behavior observed by independent observers and the behavior recorded by self-observers. A major difficulty in carrying out self-recording reliability studies is insuring the unobtrusiveness of the independent observer, who could otherwise serve as a reactive stimulus to change the reliability of self-recording, as previously mentioned. Reid (1970) reported that observers obtained median reliabilities of .75 when they were aware that reliability was being assessed, but that the median reliability dropped to .51 when they were told reliability was not being assessed. Romanczyk et al. (1973) also found that reliability was lower when observers were not informed of the reliability assessment than when observers were instructed that a reliability assessment was being performed.

Despite these difficulties in assessing the reliability of self-recorders, Simkins (1971) has held that the therapeutic potentiality of self-recording cannot adequately be investigated without further evidence of its reliability as an assessment tool. Nelson and McReynolds (1971) agreed that ideally both reliability and reactivity of self-recording should be investigated, but not necessarily in Simkins' suggested temporal sequence. They feel that a discrimination should be made between the

reliability of self-recording and the reliability of its effects. It is possible that self-recording has reactive effects on an individual's behavior even while the self-recording is highly unreliable.

Considering the above problems, Lipinski and Nelson (1974) assessed, separately, the reactivity and reliability of self-recording. Subjects, as well as independent observers, recorded face-touching behavior. The results indicated that, independent of low reliability coefficients for subjects engaged in self-recording, self-observing was indeed a reactive process. The independent observers recorded a large decrement in face-touching when the subjects were self-recording their own behavior. Not only was self-recording reactive, but the self-recorders were unreliable as compared with the independent observers. When the self-observers were aware of the reliability check, their reliability was .86, but their reliability dropped to .52 when they were unaware of the reliability check.

Statement of the Problem

The data reviewed in the present paper suggest that self-recording can serve two functions: that of a method of data collection and that of therapeutic tool. As a method of baseline and intervention data collection,

self-recording would provide a frequency count of behaviors as well as allowing for identification of antecedents and consequences. The primary importance of self-recording would be found when the target behaviors were "private." These private behaviors could include cognitions, or overt behaviors that are usually performed privately. As a therapeutic tool, self-observation of a specified behavior may reactively cause the behavior to change in a desired direction. Behavior seems to increase if it is an appropriate behavior and seems to decrease if it is an inappropriate response. This reactivity, although therapeutically useful, of course, would interfere with the utility of self-recording as an assessment device.

The present study represents a step toward analysis of certain variables that influence self-observations. Since self-recording is unreliable as compared with data from independent observers, its role in data collection is limited. In addition, the unreliability of self-recording may affect the consistency of its therapeutic effects. If this is the case, the therapeutic function of self-recording may be restricted. The question then arises whether it is possible to enhance the effectiveness of self-recording as both an assessment device and therapeutic tool, thus, increasing the effectiveness of most treatment programs by using self-monitoring as a method

of data collection and as an agent for behavior change independent of any other specific therapeutic techniques.

In order to investigate the possibility of enhancing the effectiveness of self-recording as both an assessment device and as a therapeutic tool, the subjects in the present investigation were differentially reinforced for increases in reliability of self-recorded data or differentially reinforced for decreases in the behavior that was being self-recorded. Three hypotheses were investigated. First, differentially reinforcing subjects for increments in reliability with the data recorded by independent observers would increase the subjects' reliability. Second, differentially reinforcing subjects for decrements in the behavior that was being self-recorded would result in decreases in the behavior. Third, reliability would be lower when subjects were unaware that reliability was being assessed than when they were aware of the assessment. The latter hypothesis is based on the results from studies in which higher reliability occurred when independent observers were aware of the assessment (Reid, 1970; Romanczyk et al., 1973). It was speculated that self-recording would be subject to the same phenomenon, that is, that self-recorders may produce higher reliability with independent observers when they are aware that reliability is being calculated.

The experimental design was a 2x7x9 factorial design with subjects nested in the two treatment groups and repeated across the seven experimental conditions with nine observations under each condition. The first group of self-recorders were differentially reinforced for increases in reliability and the second group were differentially reinforced for decreases in face-touching. The seven experimental conditions were: first, a baseline period consisting of initial observations by two independent observers with subjects unaware of the observers and unaware that they would later be doing self-observations; second, a baseline-aware condition, in which self-recorders knew that they were being observed and that their reliability was being assessed; third, a baseline-unaware condition in which self-recorders did not know that they were being observed nor that reliability was being assessed while they were doing self-observations; fourth, a self-recorder aware condition in which subjects knew that they would be reinforced for either high reliability or reactive data and did know that they were being observed; fifth, a self-recorder unaware condition in which the self-recorders knew that they would be reinforced for either high reliability or reactive data but did not know that they were being observed nor that reliability or reactivity was being assessed; sixth and seventh, return-to-baseline

states instituted after the subjects had ceased self-recording. Several separate analyses of variance were computed to permit separate analyses of both reliability and reactivity of self-recording. It was hypothesized that between-group comparisons would demonstrate that the subjects reinforced for increments in reliability would increase their reliability without decreasing their face-touching when compared to the decrease behavior group. In addition, it was hypothesized that statistical analyses would show that the decrease behavior group touched their face significantly less, while remaining unreliable, when compared to the reliability group.

CHAPTER II

METHOD

Subjects

Twenty undergraduate and graduate students (two males, 18 females) in two different classrooms were used as subjects. Students displaying high frequencies of face-touching were selected from all seating locations within the classrooms. The instructors teaching the two classes were aware of which students participated in the study but did not know the experimental hypotheses under investigation.

Independent Observers

Three independent observers (one female, two males) were used for all experimental conditions. Reliability between these observers was assessed, as well as reliability between the independent observers and the self-recorders. Reliability between the independent observers was obtained by designating one of the independent observers as the reliability checker who in turn simultaneously recorded the same subjects' behavior according to a pre-arranged schedule which was not revealed to the other observers. Reliability between the independent observers and the self-recorders was assessed by the independent observers and the subjects both using similar data sheets

which were comprised of 5-minute intervals and by both utilizing the same wall clock. During the "unaware" conditions, the observers were unobtrusive behind a one-way mirror in an observation room adjacent to the classroom. During the "aware" conditions, one observer (always the same observer) entered the classroom while the other two observers remained behind the one-way mirror.

Target Behavior

The target behavior was deliberately selected to be overt, rather than a covert behavior without public referents (Kanfer, 1970; Simkins, 1971). The target behavior was face-touching occurring in the classroom. Face-touching was defined as touching the face, head, hair or neck with a hand, pen, pencil, cup or cigarette. A new behavior occurred every time the subject's hand (or object) broke contact with the face. If two hands touched two different parts of the face, it counted as two behaviors. Objects were equivalent to face or hands, e.g., touching glasses with hands or touching pen to mouth.

Conditions

Baseline

During baseline, the independent observers, without the subjects' knowledge, recorded from behind the one-way

mirror. Baseline lasted until nine data points were obtained per subject; each data point consisted of the frequency of face-touching during a 5-minute interval. The subjects were recorded one at a time by the observers. The subjects were unaware of the observers and unaware that they would later participate in an experiment.

Baseline-Aware (B-A) and Baseline-Unaware (B-U)

The subjects were privately requested to participate in an experiment which would require them to accurately count face-touching behaviors. The subjects were given a training session on the self-recording of face-touching. A data sheet with a set of rules (see Appendix A) was given to each subject which described the target behavior. Each rule was explained to the subjects before they practiced recording the target behavior. The practice session proceeded as follows. First, a person modeled the target behavior and gave feedback to the subjects regarding self-recording of his own behavior. Second, the model engaged in the target behavior and the subjects recorded these behaviors for two 30-second intervals. At the end of each interval an explanation of the target behavior that occurred during that interval was given to the subjects. Subsequent to this training, the subjects were told that an observer would periodically enter the classroom to check their accuracy and that the data sheets would be in

the classroom at the beginning of each class. Questions were answered and the subjects were informed that they would receive money for participating in this study, but the conditions for receiving the money would be explained at a later date.

The purpose of the Baseline-Aware (B-A) and the Baseline-Unaware (B-U) conditions was to establish the self-recorders' baseline frequency of face-touching while self-recording. The B-A and B-U conditions were divided among each session and counter-balanced so that both conditions occurred approximately the same number of times at the beginning and end of class. When the B-A condition was in effect, one independent observer entered the classroom and sat near the subjects. The subjects were previously informed that the independent observer would enter the classroom to check their accuracy. The other independent observers continued to record from behind the one-way mirror. All independent observers observed the subjects on a pre-arranged basis and their stop watches were synchronized with the clock used by the subjects -- the clock was a standard classroom clock with a minute hand; the clock was located in the front of the room and was visible from all locations within the room.

In the B-U condition, the self-observers did not know that reliability or the frequency of their face-touches was

being assessed. The independent observers recorded unobtrusively from behind the one-way mirror. The rest of the procedure was the same for both conditions. The subjects recorded their own frequencies of face-touching during 5-minute intervals for the entire class period. The independent observers used data sheets similar to those of the subjects (see Appendix B) and simultaneously recorded the subjects' face-touching frequencies, observing only one subject per 5-minute interval. Nine data points were obtained for each subject in each of the informed and uninformed conditions. Usually, one data point was obtained for each subject during each class period for each of the informed and uninformed conditions.

As stated above, these two conditions were used for the purpose of establishing the self-recorders' baseline frequency of face-touching while self-recording. In addition, these conditions allowed an assessment of the reliability of the self-recorders. Both baseline frequency data and reliability coefficients were used as matching variables. That is, subjects in the reliability group and in the decrease face-touching group were matched in terms of their baseline face-touching frequency and reliability coefficient obtained during the B-A and B-U conditions. The mean frequency of face-touches for the reliability group was 5.87 and for the decrease face-touching group was 6.22.

The mean reliability for the reliability group was .584 and the decrease behavior group was .546. Two separate analyses of variance yielded no significant differences between the two groups on the frequency and reliability data (see Results Chapter).

Self-Recorder Aware (SR-A) and Self-Recorder Unaware (SR-U)

After completion of the previous condition, ten subjects were assigned to the reliability group and ten subjects to the decrease face-touching group. The subjects were given a set of instructions (see Appendices C and D) with an explanation of either the reliability or the face-touching criterion level for their group. The instructions indicated that they would receive $1.00 per session contingent on maintenance of the criterion level and that they would be paid for successive approximations to the criterion. Reinforcement for the self-recorders consisted of money distributed on the following basis:

Reliability Group: A reliability coefficient was computed for each self-observer in the reliability group during the B-A and B-U conditions. A correlation coefficient between the data reported by the self-observer and the mean of the independent observers was the method for calculation. This reliability estimate was utilized as the self-recorder baseline reliability. After the baseline reliability estimate was obtained, the self-recorder conditions were

instituted (both self-recorder aware, SR-A, and self-recorder unaware, SR-U). A desired criterion level of .90 reliability was set. Upon obtaining this level, each subject received $1.00 per session contingent upon maintenance of this level of reliability. Subjects were also paid for successive approximations to this criterion; the reinforcement schedule for each subject was dependent upon the difference between each individual subject's baseline measure and the .90 desired criterion level. For each subject, this difference was divided into $1.00 and a cents per 1%-increase payment rate was computed. For example, if Subject 1 had a baseline of .65, his payment rate to criterion level .90 was $1.00/.25 or 4¢ per 1% increase in reliability. Number of face-touches was irrelevant for the reliability group. During the SR-A condition, daily reliability (upon which reinforcement was based) was calculated by placing the mean of the independent observers' data, obtained during the SR-A condition, and the self-observer's data in ratio so that the smaller frequency appeared in the numerator.

Decrease Behavior Group: A baseline data score in terms of the mean number of face-touches per five-minute interval was obtained for each subject in the decrease behavior group during the B-A and B-U condition. After the baseline score was obtained, the SR-A and SR-U

conditions began. A desired criterion level of zero face-touching behaviors for each 5-minute interval observed by the independent observers was set. Upon obtaining this level, each subject received $1.00 per session contingent on maintenance of this level of reported face-touches. Subjects were also paid for successive approximations to this criterion; the reinforcement schedule for each subject was dependent upon the difference between each individual subject's baseline measure and the zero criterion level. For each subject, this difference was divided into $1.00 and a cents per one face-touching behavior decrease payment rate was computed. For example, if Subject 2 had a baseline of 10 behaviors, his payment rate to criterion level was $1.00/10 or 10¢ per each face-touching behavior decrease. Reliability scores were irrelevant for the decrease behavior group.

To allow time for data computation, the cash reward earned in the previous session was given to the subjects before the start of the next session. They were informed about the amount that their reliability had increased or that their face-touching had decreased on the previous day's session. For instance, the subjects were given an envelope at the start of each SR-A session. The envelope contained a note (see Appendix E) bearing the increase (or decrease) in reliability (or face-touching) since the

previous session in addition to their monetary reinforcement, if any, earned during the previous session. The subjects' reliability or face-touching scores were determined only on their behavior while the independent observer was in the classroom.

The remainder of the procedure was the same as previously described for the B-A and B-U conditions. The SR-A and SR-U conditions were equally divided within each session. Three independent observers recorded the data and the subjects continued to record their own frequencies of face-touching. When the SR-A condition was in effect, one independent observer entered the classroom and sat near the subjects. During the SR-U condition, the self-observers did not know that reliability or the frequency of face-touches was being assessed. Nine data points were obtained per condition. The rationale for the SR-U condition was to permit a test for generalization of the effects produced by reinforcement based on the SR-A condition.

After the SR-A and SR-U conditions were completed, the subjects were informed that the experiment was finished. As a check on the manipulation, all subjects completed a questionnaire (see Appendix F). Each subject was sent a letter after completion of the study describing the effect of self-recording on their behavior and thanking them for participating in the study.

Return-to-Baseline I and II

Although the subjects were informed that the experiment had terminated, the observers continued to record the subjects' behavior until nine data points, each consisting of the frequency of face-touches during a 5-minute interval, were obtained for each subject during each post-testing condition. Two post-testing conditions occurred since it appeared that the frequencies of face-touching were lower during Post-Testing I than during the initial baseline; an additional nine data points (Post-Testing II) therefore were collected for each subject. During these post-testing conditions, the independent observers remained unobtrusive behind the one-way mirror.

CHAPTER III

RESULTS

Reactivity

Mean daily face-touches were calculated from the observations taken by the independent observers for both the decrease behavior and increase reliability groups. The means for these groups in each of the seven experimental conditions are presented in Figure 1.

Examination of the figure suggests that both groups showed a substantial decrease in face-touching from baseline to the first self-record conditions (B-A and B-U), and that a further decrement in face-touching occurred during the reinforcement conditions. An increment in face-touching was noted when self-recording was terminated (return-to-baseline).

An analysis of variance, summarized in Table 1, substantiated the observed decrement in face-touching. As hypothesized the experimental conditions (A) were significant ($F = 92.70$; $df = 6, 108$; $p < .01$), but no significant difference between the groups (B) was observed. The magnitude of the experimental conditions effect was robust ($\omega^2 = .45$). Newman-Keuls means comparisons revealed that the mean performance during baseline was statistically different ($p < .01$) from the mean performance during the self-record (B-A, B-U, SR-A and SR-U) conditions. The

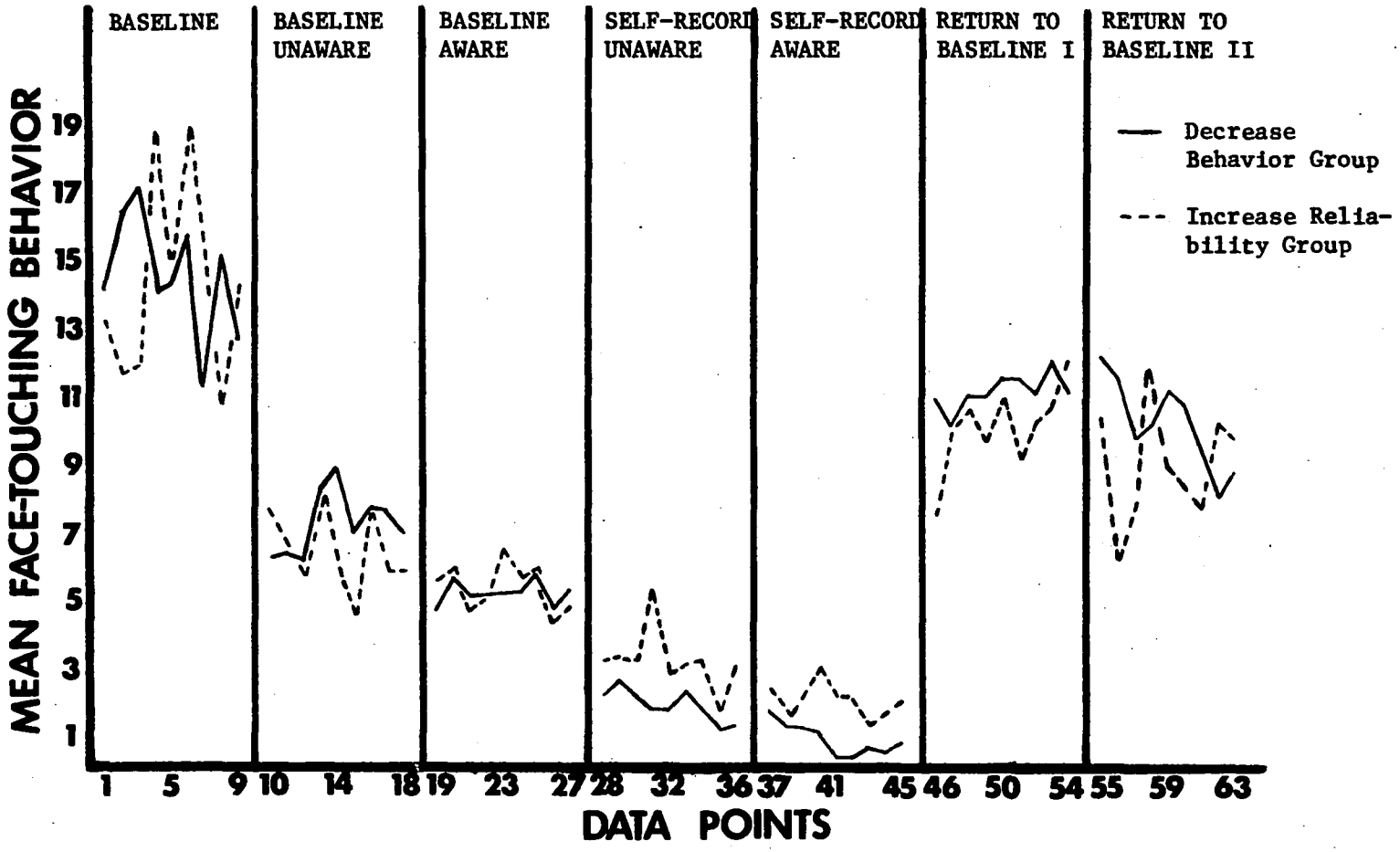

Figure 1. Mean face-touching behavior for the two groups of subjects per 5-minute interval. The unknown and known conditions were interspersed among each other but are presented separately.

Table 1

Groups (2) x Conditions (7) x Observations (9)

Repeated Measures Analysis of Variance

for Face-Touching

Source	df	MS	F
Between Subjects			
Reinforcement (B)	1	1.54	.0126
Subj. w. groups	18	121.56	
Within Subjects			
Experimental Conditions (A)	6	3852.99	92.70**
A X B	6	54.88	1.32
A X Subj. w. groups	108	41.56	
Observations (C)	8	21.49	1.17
B X C	8	26.37	1.44
C X Subj. w. groups	144	18.32	
A X C	48	13.99	.714
A X B X C	48	17.86	.911
A X C X Subj. w. groups	864	19.62	

**p < .01

mean performance for the two return-to-baseline conditions both were significantly different from the SR-U and SR-A conditions but were not statistically different from B-U and B-A. There were no other statistically significant differences at the .01 level for the Newman-Keuls test. The data indicated that face-touching showed a significant decrease during the self-record conditions which was followed by an increase after discontinuing the self-recording but this increase was not significantly different from baseline.

A groups X self-record conditions X observations (2x4x9) repeated measures analysis of variance, utilizing the independent observers' data, was performed (see Table 2) on the frequency of face-touching. Since the subjects were matched in groups on their frequency of face-touching for two (B-A, B-U) of the four conditions, it was predicted that there would be no significant group differences, only a significant difference between conditions. Confirming this prediction, the results from the analysis revealed a significant conditions effect ($F = 98.90$; $df = 3,54$; $p < .01$), a significant condition X group interaction ($F = 5.22$; $df = 3, 54$; $p < .01$), and an observation main effect ($F = 2.18$; $df = 8, 144$; $p < .05$).

The condition X group interaction was examined via a Newman-Keuls test. Only two significant differences,

Table 2

Groups (2) x Self-Record Conditions (4) x Observations (9) Repeated Measures Analysis of Variance for Face-Touching

Source	df	MS	F
Between Subjects			
Reinforcement (B)	1	46.51	.95
Subj. w. Groups	18	49.16	
Within Subjects			
Experimental Conditions (A)	3	1062.10	98.90**
A X B	3	56.07	5.22**
A X Subj. w. Groups	54	10.74	
Observations (C)	8	10.93	2.18*
B X C	8	4.33	.86
C X Subj. w. Groups	144	5.01	
A X C	24	5.34	.91
A X B X C	24	4.70	.80
A X C X Subj. w. Groups	432	5.87	

*p < .05
**p < .01

both at .05 level, appeared. The B-U condition was statistically different from both SR-U and SR-A for the decrease behavior group. It appeared that the condition X group interaction was relatively weak since the magnitude of the effect was low ($\underline{w}^2$ = .016). Figure 2 is a geometric representation of the interaction.

A large proportion of variance (37%) was accounted for by the self-record conditions main effect. The Newman-Keuls procedure was used to test the differences between all possible pairs of means. The mean face-touches during the B-U condition was significantly higher than the mean face-touches for SR-U and SR-A. The mean performance for the B-A condition was significantly higher (at .01) than the mean face-touches for SR-A with the difference between B-A and SR-U significant at the .05 level. The above data demonstrated that self-recording was reactive, independent of the reliability of the self-recorders, and that self-observers reduced their frequency of face-touching when they were reinforced for either increasing their reliability or reducing their face-touches.

Since the two treatment groups were matched on their face-touching from the B-A and B-U conditions, a 2x2x9 (groups X baseline self-record conditions X observations) repeated measures analysis of variance was done on the frequency of face-touching during B-A and B-U. The

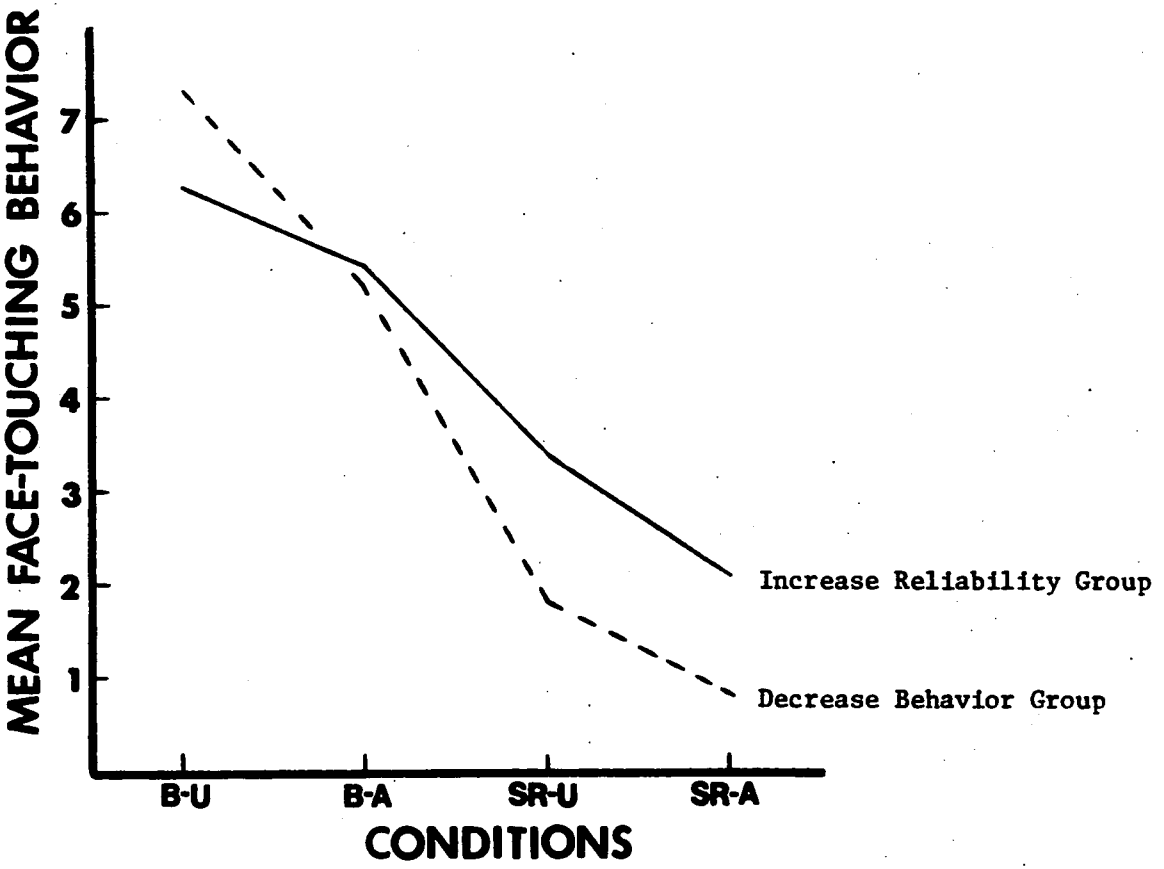

Figure 2. Mean face-touching behavior for the two groups of subjects per 5-minute interval. The A X B interaction is depicted.

analysis denoted successful matching since there was no significant difference between the groups (see Table 3). The same analysis revealed that there was a significant main effect for conditions (F = 11.79; df = 1, 18; $p < .01$). The amount of variance accounted for by this main effect was small (w^2 = .044).

Table 4 summarizes the groups X reinforcement self-record conditions X observations (2x2x9) repeated measures analysis of variance for the SR-U and SR-A conditions which indicated that the self-recorders reinforced for decreasing their face-touching reduced the frequency of their face-touches significantly more than self-recorders reinforced for increasing their reliability (F = 10.83; df = 1, 18, $p < .01$). Therefore, the hypothesis that differentially reinforcing subjects for decrements in the behavior that is being recorded will result in decreased face-touching was supported by the present analysis. The magnitude of the above main effect was w^2 = .11. The analysis also showed that significantly fewer face-touches occurred during the SR-A than during the SR-U condition (F = 41.36; df = 1, 18; $p < .01$). However, the magnitude of this effect (w^2) was only .076.

Reliability for Self-Recorders

The reliability of the self-observers was calculated for each of the four self-record conditions. When

Table 3

Groups (2) x Baseline Self-Record Conditions (2) x Observations (9) Repeated Measures Analysis of Variance for Face-Touching

Source	df	MS	F
Between Subjects			
Reinforcement (B)	1	11.38	.248
Subj. w. Groups	18	45.96	
Within Subjects			
Experimental Conditions (A)	1	202.50	11.79**
A X B	1	31.21	1.82
A X Subj. w. Groups	18	17.17	
Observations (C)	8	11.51	1.25
B X C	8	4.80	.50
C X Subj. w. Groups	18	9.55	
A X C	8	7.21	.84
A X B X C	8	6.67	.78
A X C X Subj. w. Groups	144	9.54	

**p < .01

Table 4

Groups (2) x Reinforcement Self-Record Conditions (2) x Observations (9) Repeated Measures Analysis of Variance for Face-Touching

Source	df	MS	F
Between Subjects			
Reinforcement (B)	1	169.47	10.83**
Subj. w. Groups	18	15.64	
Within Subjects			
Experimental Conditions (A)	1	107.80	41.36**
A X B	1	2.67	1.02
A X Subj. w. Groups	18	2.61	
Observations (C)	8	6.40	2.84*
B X C	8	5.03	2.24*
C X Subj. w. Groups	144	2.25	
A X C	8	1.84	.81
A X B X C	8	1.95	.85
A X C X Subj. w. Groups	144	2.27	

*p < .05
**p < .01

reliability was being taken by the independent observers recording the behavior of the same subject, a mean was computed, and this mean was compared with the self-observers' values. A Spearman coefficient was used for calculating the reliability coefficient for every subject during each condition (see Table 5). Spearman correlation coefficients were used rather than Kendall's tau or Pearson because they appeared to represent the data more appropriately. According to Nie, Bent, and Hull (1970), there is no fixed rule for selecting one method over the other; in fact, the concepts underlying the Spearman and Kendall coefficients are similar and usually yield similar results when computed on the same data. As for the Pearson, Nie et al. (1970) suggested that it should only be used with interval scales, and that Spearman or Kendall should be used with ordinal variables (the present study utilized ordinal data). However, as indicated above, there is no set rule for using one correlation coefficient over the other, or for using Pearson only with interval data.

A groups X self-record conditions (2x4) repeated measures analysis of variance on reliability coefficients during self-recording revealed a significant main effect for conditions ($F = 8.25$; $df = 3, 54$; $p < .01$). The magnitude of the effect (w^2) was .198. As expected, no other

Table 5

Spearman Coefficient for Every Subject During Each Condition

Decrease Behavior Group	Base-U	Base-A	SR-U	SR-A
S-1	.577	.891	.892	.619
S-2	-.729	.974	1.000	1.000
S-3	.881	.773	.891	.980
S-4	.371	.009	.844	.355
S-5	.995	.945	.503	.950
S-6	.961	.897	.844	.800
S-7	.566	.604	.774	.966
S-8	.153	.238	.342	.369
S-9	-.066	.915	.711	.846
S-10	.264	.704	.793	.587
MEAN	.397	.695	.759	.747
Increase Reliability Group	**Base-U**	**Base-A**	**SR-U**	**SR-A**
S-11	.905	.527	.788	.921
S-12	.085	.879	.679	.969
S-13	.051	.353	.908	.974
S-14	.328	.819	.808	.929
S-15	.621	.136	.923	.995
S-16	.145	.861	.940	.951
S-17	.724	.848	.939	.947
S-18	.871	.771	.869	1.000
S-19	.922	.898	.895	1.000
S-20	.427	.468	.745	1.000
MEAN	.512	.656	.849	.969

significant differences were observed (see Table 6). The reason for no significant group effects was the fact that the groups were matched on reliability coefficients for two of the four conditions.

A Newman-Keuls test on the means from the four experimental conditions revealed three significant differences. The mean reliability for B-U (.46) was statistically different from the mean reliability for all other conditions (B-A, .67; SR-U, .81; SR-A, .86). The hypothesis that reliability would be lower when subjects were unaware that reliability was being assessed than when they were aware of the assessment thus received only partial support from the data analysis. That is, the reliability from B-U was significantly lower than the self-recorders' reliability during B-A but the reliability coefficients between SR-U and SR-A were not significantly different. The latter results may possibly be explained by a generalization of the effects of reinforcement for increased reliability during SR-A and SR-U.

Figures 3 and 4 represent a comparison of the data collected by the self-recorders with the data collected by the independent observers for the decrease behavior group (Figure 3) and the increase reliability group (Figure 4). The data (which were averaged) were the means of the observations made during each session. These

Table 6

Groups (2) x Self-Record Conditions (4) Repeated Measures Analysis of Variance for Reliability Coefficients

Source	df	MS	F
Between Subjects			
Reinforcement (B)	1	.188	1.70
B X Subj. w. Groups	18	.110	
Within Subjects			
Experimental Conditions (A)	3	.643	8.25**
A X B	3	.057	.73
A X Subj. w. Groups	54	.078	

**p < .01

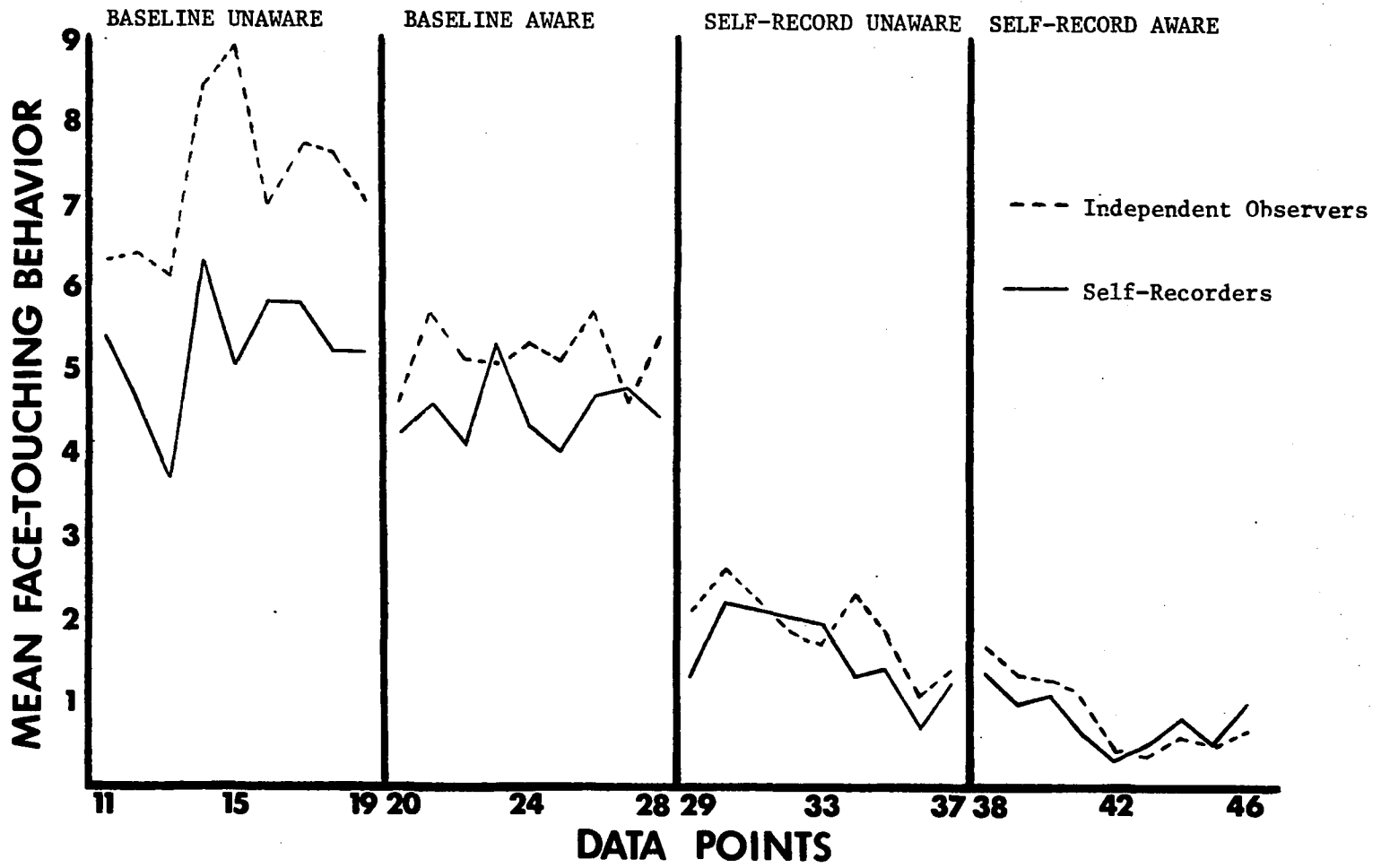

Figure 3. Mean face-touching behavior recorded by the independent observers and self-recorders in the decrease behavior group per 5-minute interval. The unknown and known conditions were interspersed among each other but are presented separately.

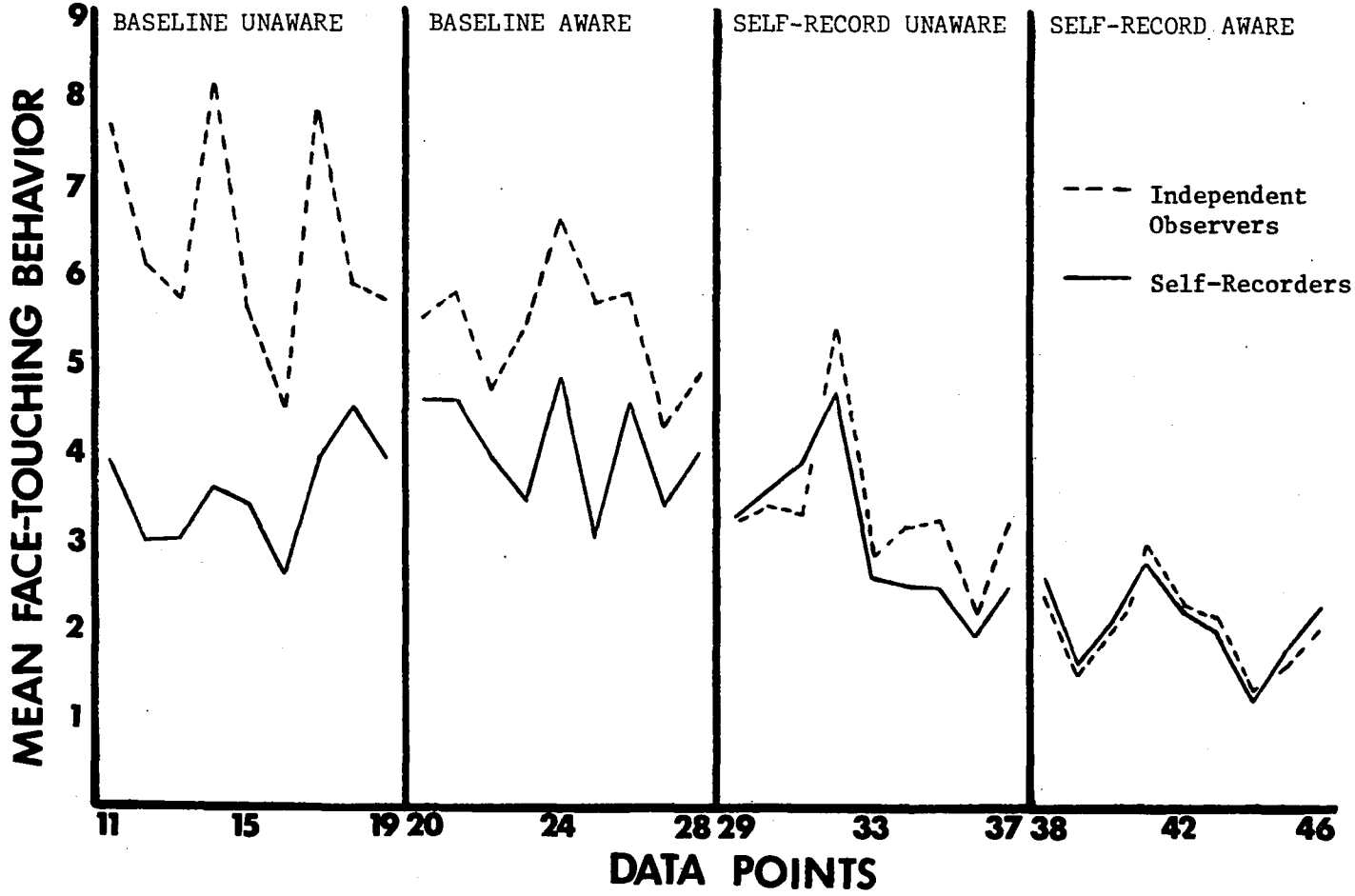

Figure 4. Mean face-touching behavior recorded by the independent observers and self-recorders in the increase reliability group per 5-minute interval. The unknown and known conditions were interspersed among each other but are presented separately.

figures give an indication of the agreement between the independent observers and self-recorders during the reinforcement conditions; the congruence of the data in Figure 4 between the self-recorders who were being reinforced to increase reliability and the independent observers in SR-A should especially be noted. The figures also display the vast differences between the data collected by the independent observers and that reported by the self-observers, during the B-U condition, thus displaying the low reliability for this condition.

Since the subjects were also matched to groups according to their reliability scores during the B-U and B-A conditions, two separate 2x2 analyses of variance were done on the reliability data. In the first analysis, B-U expectedly failed to yield any significant differences; thus, showing that the groups were equally matched (see Table 7). The second analysis, SR-U X SR-A, revealed a statistically significant difference ($\underline{F} = 6.80$; $\underline{df} = 1, 18$; $\underline{p} < .05$) between the self-recorders reinforced for increasing their reliability and the self-recorders reinforced for decreasing their behavior (see Table 8). The magnitude of the present effect ($\underline{w}^2$) was .154. The mean coefficient for the self-recorders reinforced for increasing their reliability was .909 and the group reinforced for decreasing their behavior had a coefficient

Table 7

Groups (2) x Baseline Self-Record Conditions (2)

Repeated Measures Analysis of Variance

for Reliability Coefficients

Source	df	MS	F
Between Subjects			
Reinforcement (B)	1	.015	.098
B X Subj. w. Groups	18	.148	
Within Groups			
Experimental Conditions (A)	1	.487	3.44
A X B	1	.059	.42
A X Subj. w. Groups	18	.142	

Table 8

Groups (2) x Reinforcement Self-Record Conditions (2) Repeated Measures Analysis of Variance for Reliability Coefficients

Source	df	MS	F
Between Subjects			
Reinforcement (B)	1	.242	6.80*
B X Subj. w. Groups	18	.035	
Within Groups			
Experimental Conditions (A)	1	.029	1.48
A X B	1	.043	2.23
A X Subj. w. Groups	18	.020	

*p < .05

of .753. Therefore, the hypothesis that differentially reinforcing subjects for increments in reliability would increase the subjects' reliability was supported by the present analysis.

Reliability for Independent Observers

The reliability coefficients between the independent observers were calculated separately for each condition. A Spearman coefficient which had previously been used to calculate the reliability between the independent observers and the self-recorders was also used for calculating the reliability for the independent observers (see previous explanation for using the Spearman coefficient). The data used were frequency counts per 5-minute recording intervals. The correlations, population size, and significance levels are found in Table 9. The over-all reliability between the independent observers was .976, $p < .001$.

Check on Manipulation

After the subjects had completed self-recording, they each completed a questionnaire (see Appendix F). One of the questions (#4) asked the subjects to indicate how the accuracy of their self-recording changed during the study. The majority of the subjects stated that they either noted no change or became more accurate during the

Table 9

Spearman Coefficients for the Independent Observers During Each Condition

Condition	N	Spearman	Significance Level
Baseline	68	.95	$p < .001$
B-A	74	.91	$p < .001$
B-U	72	.94	$p < .001$
SR-A	70	.96	$p < .001$
SR-U	66	.93	$p < .001$
Post-Testing I	72	.96	$p < .001$
Post-Testing II	72	.97	$p < .001$
Over-all Reliability	494	.97	$p < .001$

study. A *t* test for independent groups failed to reveal a significant difference between the groups in terms of their responses to the question. Only two of the 20 subjects felt that their accuracy may have been affected by a lack of understanding of the rules for self-recording (question #5).

The self-recorders were also asked how self-recording affected the frequency of their face-touching (question #6). The subjects in the decrease behavior group all specified that they either noted a slight or great decrease in the frequency of face-touching, whereas the response from the group reinforced for increments in reliability varied, from a great decrease to no change to a slight increase in the frequency of face-touching. A *t* test for independent samples failed, however, to yield a significant difference between the groups on their responses to question 6.

As a check on the reinforcement manipulation, the subjects were asked why they were receiving reinforcement (question #10). All of the subjects in the decrease behavior group, as well as in the increase reliability group, knew why they were receiving the reinforcement. Most of these subjects indicated that the reinforcement was an incentive for obtaining the criterion level (question #11). In order to assess the unobtrusiveness of the independent observers behind the one-way mirror, question 9 asked,

"What aspects of this study, if any, aroused your suspicion?" One of the 20 subjects specifically commented on the one-way mirror but the subject's comment pertained to using the one-way mirror in a future study. None of the subjects indicated that they believed any one was observing them from behind the one-way mirror.

CHAPTER IV

DISCUSSION

Results from the present study showed that the subjects decreased their face-touching when they were self-recording. A large decrement in face-touching was recorded by the independent observers when the subjects were self-recording. The mean face-touching during baseline was 14.49, while the mean dropped to 6.04 (averaged mean of both conditions) during the B-U and B-A conditions. A further reduction in face-touching was noted during the reinforcement conditions. These findings corroborated previous reports that self-recording produces behavior change. McFall (1970) found that self-recording of in-class smoking increased its frequency, while self-observations of in-class abstinence decreased the frequency of smoking. McFall and Hammen (1971) reported that smokers decreased their frequency of smoking by self-recording. Johnson and White (1971) showed that self-observing studying behavior increased the students' grades. In addition, Broden et al. (1971), Ernst (1973), Gottman and McFall (1972), and Rutner and Bugle (1969) affirmed that self-recording is reactive to the extent that it changes behavior in a desired direction. The decrease in face-touching noted in the present study supported previous studies that utilized face-touching as the target behavior (Lipinski & Nelson, 1974; Nelson et al., 1974).

Kanfer (1970) has hypothesized that a feedback loop, e.g., the alteration of behavior as a function of response feedback, may be a possible mechanism operating during self-recording that permits a subject to adjust his own behavior. Many of the subjects in the present experiment appeared to be aware of the decrement that occurred in their face-touching. When asked in the post-self-recording questionnaire how self-recording affected the frequency of their face-touching, 18 out of 20 subjects reported either a slight or great decrease in their face-touching behavior, when in fact a decrease had been recorded by the independent observers. Previous studies (Lipinski & Nelson, 1974; Nelson et al., 1974) have also found that subjects were aware of the effects of self-recording on their behavior.

The results from the present study also confirmed that self-recorders are unreliable, especially when they are unaware that reliability is being assessed. The reliability for the self-recorders during the B-U condition was .46 compared to .67 for B-A. McFall (1970), without manipulating awareness of reliability assessment, reported a correlation of .61 between his self-observers and independent observers. Lipinski and Nelson (1974) found that the reliability between self-recorders and independent observers dropped from .86 to .52 when they were unaware that

reliability was being evaluated. In a later study, Nelson et al. (1974) also showed that self-recorders are more reliable when they are aware of the reliability assessment. When the self-recorders were aware that reliability was being assessed, their reliability coefficient was .810, but their reliability dropped to .554 when they were unaware of the assessment. These findings about self-recorders parallel those about independent observers. Reid (1970) reported that observers obtained median reliabilities of .75 when they were aware that reliability was being assessed, but that the median reliability dropped to .51 when they were told that reliability would not be assessed. Romanczyk et al. (1973) found that reliability was lower when observers were not informed of the reliability assessment than when the observers were instructed of the reliability assessment.

It would seem that reliability of self-recorded data would be necessary if self-recording is to be utilized as a method of data collection. Lack of consistent recording would give the self-recorder, as well as the therapist, false feedback on the frequency of the target behavior. It is likely that the target behavior would be either under- or over-estimated. It is possible, however, that highly unreliable self-recording may nonetheless alter the frequency of the target behavior (Nelson & McReynolds, 1971).

When self-recording is to be used primarily as a therapeutic tool, some other more reliable assessment device must also be used in order to accurately assess the reactive effects of self-recording, and to provide feedback for both the self-observers and therapist. Since the issue of reliability of self-recording is important (Nelson & McReynolds, 1971; Simkins, 1971), attempts to improve the reliability of self-recorders or to obtain independent behavior ratings should be made.

The primary purpose of the present investigation was to enhance the effectiveness of self-recording as both a method of data collection and as a therapeutic tool by differentially reinforcing self-recorders for increasing their reliability or for decreasing their target behavior. The results indicated that subjects reinforced for increasing their reliability showed an increment in reliability over baseline levels as well as over the decrease behavior group. These data suggest that the self-recorders' reliability can be increased by reinforcing such increments. The fact that reinforcement contributed to the effect was exemplified by the subjects' response that reinforcement was an incentive for obtaining the criterion level. Additional research is needed using other target behaviors in different environmental settings. The present data do imply that the reliability of self-recorders can be

enhanced so that self-recording can be used as a method of data collection.

The self-recorders who were reinforced for decreasing their face-touching behavior reduced their face-touches significantly more than the group reinforced for improving their reliability. The results from the present study supported Kanfer's (1970) assumption that reinforcement is an important variable in self-recording that helps produce behavior change. The data suggest that further reductions in the target behavior may be possible by reinforcing the subjects for such decrements. The importance played by reinforcement in reducing face-touching was noted in the subjects' awareness of why they were receiving reinforcement, and their indication of the incentive that reinforcement provided for reducing their face-touches. Again, additional data are needed on other target behaviors to determine if reinforcement affects them in similar ways. If so, it would be possible to use self-recording in conjunction with reinforcement to bring about faster and greater changes in the response being self-recorded.

These data support Nelson and McReynolds' (1971) assumption that the reliability and reactivity of self-recording procedures could be dealt with as separate issues. During the B-U and B-A conditions, self-recording had reactive effects on the subject's face-touching, even

while they were not accurate when compared with independent observers. When the subjects were reinforced for increasing their reliability, they did so without reducing their behavior significantly more than the group reinforced for decreasing their behavior (face-touches). The self-recorders reinforced for decreasing their face-touches reduced this behavior while remaining unreliable.

The effectiveness of self-recording in reducing face-touching behaviors was only temporary. When self-recording was discontinued, a subsequent increment in face-touching was recorded by the independent observers. Although there was an increase in face-touching behavior, it did not reach the prior baseline frequency. An additional nine data points still showed that the subjects were below their baseline data; the differences between baseline and return-to-baseline conditions were not significant. The baseline mean was 14.49 compared to 10.15 which was the mean (both conditions averaged together) for the return-to-baseline conditions. These results contradicted the findings of McFall (1970) who reported that the decrement in smoking behavior that occurred during self-monitoring was maintained during a follow-up period, the results of McFall and Hammen (1971) who found that reductions in smoking behavior were maintained following self-recording, and the data of Gottman and McFall (1972) who found a high rate of

talking after self-recording ceased. However, Broden et al. (1971) achieved a return-to-baseline for studying and talking when self-recording was no longer in effect. Similarly, Lipinski and Nelson (1974) and Nelson et al. (1974) showed that face-touching increased to baseline levels after discontinuing self-recording. Maletzky (1974) also reported that response frequencies increased when his clients stopped self-observing, but decreased when they again self-recorded their target behavior.

The discrepant data on the maintenance of behavior change after the discontinuation of self-recording may be a function of several variables. First, the subjects' (clients') motivation for change can be an important factor in determining the frequency of the target behavior after the cessation of self-recording. The subjects in McFall and Hammen's study all indicated that they wanted to stop smoking. The subjects in Broden et al. (1971), Lipinski and Nelson (1974), and Nelson et al. (1974) studies did not specifically express a desire to alter the frequency of their target behavior. Thus the subject's motivation may be an important variable that needs further study. Next, the target behavior being self-recorded and the length of self-recording may be important factors in determining any maintenance of change in the target behavior after stopping self-recording. Certain target

behaviors may have a longer history of reinforcement and hence be more stable than others, and may therefore require longer periods of self-recording. For example, a person with a long history of smoking would probably have to self-record for weeks or months before he acquires self-control over his smoking. Terminating self-recording before a person has obtained self-control would result in a return to the base rate. Closely related to the above point is the value orientation of the behavior. Assuming that self-recording alters the person's awareness of the response class that is being self-recorded, the value judgment attached to the behavior may set the occasion for self-evalution. This self-evaluation, Kanfer (1970) has speculated, would be self-reinforcement for behavior changes which are perceived as desirable. Therefore, self-reinforcement would maintain the desired behavior after stopping self-recording. It is possible that self-reinforcement or even external reinforcement may account for the longevity of reactive effects. Clearly, additional data are needed on the variables that effect the long-term consequences of self-recording.

The implications of the above findings are numerous. First, self-recording can be used as a method of data collection. Since self-recording is likely, however, to be unreliable, every attempt should be made to obtain

periodic independent behavior ratings. This can be accomplished by having independent observers, e.g., spouse or friend, occasionally check the accuracy of the self-observers. The reliability of self-recording can possibly be enhanced by having these independent observers reinforce self-recorders for increments in their reliability. Self-recording is especially important as an assessment device for coverants (Homme, 1965), such as urges, thoughts, and images, because such covert events are not observable to independent observers. The major problem with using self-recording for coverants is this lack of an adequate means for obtaining independent observations.

Second, it appears that self-recording is a reactive technique that can be used to therapeutically increase appropriate behavior or to decrease inappropriate behavior. The limitations on the kinds of behaviors for which self-monitoring may be used have not been determined; but reactive effects have been shown in smoking, studying, disruptive behavior, mouth biting, face-touching, talking, and hallucinatory behavior. The reactive effects of self-observing may be heightened by reinforcing the self-recorders for further changes in the target behavior. Despite lack of reliability, the self-observers in the present study greatly decreased their face-touching when reinforced for such decreases. It is possible that other target behaviors could be influenced in similar ways.

Finally, the temporary effects of self-monitoring noted in the present study and in previous studies suggest that such variables as the strength of the target behavior, the duration of self-recording, the nature of the target behavior, and the subjects' motivation for change require further investigation. While some studies have collected immediate follow-up data, no researchers have attempted to examine the long-term effects of self-recording.

In summary, the present data have shown that the self-recording of face-touching is a reactive process that can be heightened by using reinforcement. Self-monitoring was also found to be unreliable, and influenced by the presence or absence of an independent observer. In addition, the reliability of self-observers can be increased by reinforcing them for such increments. The major finding of the present study was that the two functions of self-recording, the assessment and therapeutic functions, could be separated and differentially reinforced. If other relevant variables affecting the direction, stability, and magnitude of the reactive changes in the target behavior can be determined, it would be possible to use self-monitoring to fit the therapeutic goals of many individuals. Eventually self-observation would give these individuals the skill to adjust their own behavior without continued dependence on environmental control.

BIBLIOGRAPHY

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

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

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

Ernst, F. Self-recording and counterconditioning of a self-mutilative compulsion. Behavior Therapy, 1973 4, 144-146.

Goldfried, M. R. and Kent, R. N. Traditional vs. behavioral personality assessment: A comparison of methodological and theoretical assumptions. Psychological Bulletin, 1972, 77, 409-420.

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.

Homme, L. E. Perspectives in psychology: XXIV. Control of coverants, the operants of the mind. The Psychological Record, 1965 4, 501-511.

Johnson, S. M. and Bolstad, O. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, and E. J. Mash [Eds.], Behavior change: Methodology concepts and practice. Champaign, Illinois: Research Press, 1973.

Johnson, S. M. and Lobitz, G. K. Demand characteristics in naturalistic observation. Unpublished manuscript, University of Oregon, 1972.

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.

Kass, R. E. and O'Leary, K. D. The effects of observer bias in field-experimental settings. Symposium presented at the University of Kansas, Lawrence, Kansas, 1970.

Kent, R. N., O'Leary, K. D., Diament, C., and Dietz, A. Expectation biases in observational evaluation of therapeutic change. Unpublished manuscript, State University of New York at Stony Brook, 1973.

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

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

Mash, E. J. and McElwee, J. D. Situational effects on observer accuracy: Behavioral predictability, prior experience, and complexity of coding categories. Child Development, in press.

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.

McNamara, J. R. and MacDonough, T. S. Some methodological considerations in the design and implementation of behavior therapy research. Behavior Therapy, 1972, 3, 361-378.

Mischel, W. Personality and assessment. New York: Wiley, 1968.

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

Nelson, R. O., Lipinski, D. P., and Black, J. L. Manipulation of the direction of reactive changes in self-recording. Unpublished manuscript, University of North Carolina at Greensboro, 1974.

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

O'Leary, K. D. The effects of observer bias in field-experimental settings. Unpublished manuscript, State University of New York at Stony Brook, 1973.

O'Leary, K. D., Kent, R. N., and Kanowitz, J. Shaping data consonant with experimental hypotheses. Unpublished manuscript, State University of New York at Stony Brook, 1973.

O'Leary, K. D., Romanczyk, R. G., Kass, R. E., Dietz, A., and Santogrossi, D. Procedures for classroom observation of teachers and children. Unpublished manuscript, State University of New York at Stony Brook, 1971.

Patterson, G. R. and Cobb, J. A. A dyadic analysis of "aggressive" behavior. In J. P. Hill [Ed.], Minnesota symposium on child psychology. Volume 5. Minneapolis: University of Minnesota Press, 1971. Pp. 72-129.

Patterson, G. R. and Gullion, M. E. Living with children: New methods for parents and teachers. Champaign, Illinois: Research Press, 1968.

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

Polansky, N., Freeman, W., Horowitz, M., Irwin, L., Papanis, N., 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.

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.

Rosenthal, R. On the social psychology of the psychological experiment: The experimental hypothesis as unintended determinant of experimental results. American Scientist, 1963, 51, 268-283.

Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.

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.

Scott, P., Burton, R. V., and Yarrow, M. Social reinforcement under natural conditions. Child Development, 1967, 38, 53-63.

Sidman, M. Tactics of scientific research. New York: Basic Books, 1960.

Simkins, L. The basis of psychology as a behavioral science. Waltham, Massachusetts: Blaisdell, 1969.

Simkins, L. The reliability of self-recorded behavior. Behavior Therapy, 1971, 2, 73-87.

Skindrud, K. Field evaluation of observer bias under overt and covert monitoring. In L. A. Hammerlynck, L. C. Handy, and E. J. Mash [Eds.], Behavior change: Methodology, concepts and practice. Champaign, Illinois: Research Press, 1973.

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

White, G. D. Effects of observer presence on mother and child behavior. Cited in L. A. Hammerlynck, L. C. Handy, and E. J. Mash [Eds.], Behavior change: Methodology, concepts and practice. Champaign, Illinois: Research Press, 1973.

Wright, H. F. Recording and analyzing child behavior: With ecological data from a town. New York: Harper and Row, 1967.

Appendix A

Data Sheets Used by Self-Recorders

Name _____ Date _____

Time Interval (5-minute blocks)	Number of times face, head, hair, or neck touched
8:10 - 8:14	
8:15 - 8:19	
8:20 - 8:24	
8:25 - 8:29	
8:30 - 8:34	
8:35 - 8:39	
8:40 - 8:44	
8:45 - 8:49	

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 B

Data Sheet Used by Independent Observers

Name of Observer _____

Date _____

	5-Minute Intervals							
	1	2	3	4	5	6	7	8
Time _____ Subject ___								
Time _____ Subject ___								
Time _____ Subject ___								
Time _____ Subject ___								
Time _____ Subject ___								
Time _____ Subject ___								
Time _____ Subject ___								
Time _____ Subject ___								

Appendix C

Instructions Given to Increase Reliability Subjects

As I have previously indicated, you will be paid for participating in this experiment. However, the amount of money that you will earn depends on your reliability. Reliability simply means how accurate you are recording your face-touches compared to another observer. I will be coming into each class for approximately half the class period for the next few weeks in order to obtain a random sample of your reliability.

A desired criterion level of .90 reliability has been set. You will receive $1.00 per session contingent on maintenance of this level of reliability. You will also be paid for approaching this criterion. Your payment schedule will be dependent upon the difference between your baseline reliability, which I have determined during the last few weeks, and the .90 criterion level. This difference will be divided into $1.00 and a cents per one per cent increase payment rate will be computed. For example, if your baseline measure was .65, your payment rate to criterion level .90 would be $1.00/.25 or 4¢ per one per cent increase in reliability. Remember, once you obtain the .90 level and stay above it during each class, you will receive $1.00.

Please ask me if you have any questions on the above procedure.

Appendix D

Instructions Given to Decrease Behavior Subjects

As I have previously indicated, you will be paid for participating in this experiment. However, the amount of money that you will earn depends on your obtaining and maintaining a desired criterion-level of zero face-touching behaviors per each 5-minute interval observed by me. You will receive $1.00 per session contingent on maintenance of this level. You will also be paid for approaching this criterion. The payment schedule will be dependent upon the difference between your baseline face-touching behavior, which I have determined during the past few weeks, and the zero criterion level. This difference will be divided into $1.00 and a cents per one face-touching behavior decrease payment rate will be computed. For example, if your baseline is 10 behaviors, your payment rate to criterion would be $1.00/10 or 10¢ per each face-touching behavior decrease. Remember, maintenance of this zero criterion will result in $1.00 for each class. In order to obtain a random sample of your behavior, I will come into class for approximately half of the class period for the next few weeks.

Please feel free to ask me any questions you may have on the above procedures.

At the beginning of each class you will receive an envelope containing your performance during the last class as well as the money you have earned (if any).

Appendix E

Example of Feedback Information Given to the Subjects During the Self-Recorder Aware and Unaware Conditions

Reliability Group

Date _____

Your reliability for the last class was _____. This is above (or below) the desired criterion level of .90. The amount of money you have earned is _____ because there was an increase,(decrease), (no change) of _____ from the last class meeting.

Decrease Behavior Group

Date _____

Your total number of face-touching behaviors while I was observing you during the last class meeting was _____. This is (or is not) at the desired criterion level of zero. The amount of money you have earned is _____ because there was an increase, (decrease), (no change) of _____ from the last class meeting.

Appendix F

Post-Treatment Questionnaire

Name _____
Campus 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. As this project progressed, how do you think the accuracy of your self-recording changed?

1	2	3	4	5
became much less accurate	less	no change	more	became much more accurate

5. Do you think that your accuracy was affected by your lack of understanding of the rules for self-recording listed on the bottom of the data sheet?

____ Yes

____ No

6. As you were self-recording, how do you think that self-recording affected the frequency of your face-touching?

____ no change in frequency

____ slight decrease in frequency

____ great decrease in frequency

____ slight increase in frequency

____ great increase in frequency

Appendix F (continued)

7. How certain are you that this happened?

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

8. Describe any difficulties which you had in self-recording.

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

10. Over the past few weeks you have been receiving up to a dollar during each class -- why were you receiving this money?

11. How much of an incentive was the money you received for obtaining the criterion level?

1	2	3	4
no incentive	a little incentive	some incentive	a lot of incentive

12. How often did you immediately record your face-touching after you touched yourself?

1	2	3	4
never	sometimes	most of the time	always

13. Comments, complaints, or suggestions for improvements.