*METHODOLOGICAL AND ASSESSMENT CONSIDERATIONS IN
EVALUATING REINFORCEMENT PROGRAMS
IN APPLIED SETTINGS¹*

ALAN E. KAZDIN

THE PENNSYLVANIA STATE UNIVERSITY

The extensive use of reinforcement programs in applied settings has led to experimentation that often fails to consider potential problems in design. The logic of the within-subject design is reviewed and specific designs employed in reinforcement programs are discussed. For each design (ABAB, or multiple-baseline design across behaviors, individuals, or situations), effects are discussed that make that design less powerful with respect to demonstrating the effect of the experimental variable. Problems in interpreting results of experiments in this area of inquiry are evaluated from the standpoint of internal and external validity. The issue of control groups is presented with considerations as to situations that require their use. Finally, the assessment strategy for evaluating operant programs is discussed and recommendations are made for measurement of behaviors in addition to the target response.

The application of reinforcement systems to various populations in treatment and educational settings has proliferated in recent years (Bandura, 1969; Kazdin and Bootzin, 1972; O'Leary and Drabman, 1971). In spite of the apparent success of programs applying contingent social and/or token reinforcement, the evaluation of such programs, in many instances, has revealed a failure to recognize certain methodological factors that may influence the results of experiments or their interpretation. The present paper attempts to discuss certain issues of experimental design that need to be considered in evaluating behavior modification (operant conditioning) programs. The discussion is restricted in general to those studies specifically evaluating reinforcement programs because they usually employ a within-subject experimental design.

Studies have evaluated reinforcement programs in a variety of settings, including psychi-

atric hospitals (Ayllon and Azrin, 1965) classrooms (Wolf, Giles, and Hall, 1968), sheltered workshops (Zimmerman, Stuckey, Garlick, and Miller, 1969), institutions (Burchard, 1967), home-style treatment facilities (Phillips, 1968) the home (Wahler, 1969), and several others. The design employed in such studies is referred to as the intrasubject replication design (Sidman, 1960). A brief conspectus of the rationale behind this design will permit background for presenting methodological issues. The reader is referred to several treatises with excellent descriptions of design in this area of inquiry (Baer, 1968, 1971; Baer, Wolf, and Risley, 1968; Bijou, Peterson, and Ault, 1968; Bijou, Peterson, Harris, Allen, and Johnston, 1969; McNamara and MacDonough, 1972; Risley, 1970; Risley and Baer, 1973; Risley and Wolf, 1972; Sidman, 1960; Thoresen, 1972; Wolf and Risley, 1971).

Briefly, the basic logic of the design is to determine operations that relate functionally to the performance of behavior. The effect of a variable (e.g., contingent praise) on behavior is demonstrated by the consecutive presentation, removal, and representation of the variable to a subject. Control over a behavior is demonstrated

¹The author gratefully acknowledges K. Daniel O'Leary for his reading of the manuscript. Monographs of this article are available for \$1.00 from the Business Office of the *Journal of Applied Behavior Analysis*, Department of Human Development, University of Kansas, Lawrence, Kansas 66044. Ask for Monograph #3.

if the behavior can be altered at will by altering the experimental operations. This research strategy is contrasted with the between-group approach, which seeks to demonstrate group differences after manipulation of the independent variable(s), usually in a single session. In this design, the data are subjected to statistical evaluation, and the focus is on mean differences instead of the behavior of individual subjects.

In the within-subject design, the effects of the experimental variables may be immediately observed on response rates. Under such circumstances, it is recommended to examine variables in an "improvised and rapidly changing design" (Skinner, 1969, p. 112), with the goal of achieving control over behavior. This approach bypasses variability due to intersubject differences which is included in the design of between-group experiments. This is a desirable feature from several standpoints. First, intersubject variability is not a feature of the behavioral process of the individual subject, but is an effect due in part to the method of study. Second, in evaluating experimental findings in traditional group designs, intersubject variability serves as a base for statistical evaluation of the results. Because of this, as Sidman (1960) noted, lawful effects of variables may be obscured. A similar problem is that averages from group data usually have no analogue in representing the behavioral process of individuals. The form or shape of the function obtained with group data does not necessarily represent the behavior change process of the individual. Several subjects in the group may be affected differently by the experimental manipulation. This is obscured in the between-group analysis. Having outlined, briefly, the rationale for the within-group approach, types of these designs used and their respective sources of problems may be identified.

Specific Designs

The first design is frequently referred to as the ABAB design (where A refers to baseline conditions, and B refers to the experimental con-

dition). Other names for the design include reversal technique (Baer *et al.*, 1968), intrasubject replication design (Sidman, 1960), equivalent time-samples (Campbell and Stanley, 1963), and intensive design (Chassan, 1967). The design employs alternate presentations of the baseline and experimental conditions within a subject or group of subjects. Several variations may be used in this design. More than one experimental condition may be presented before the second baseline (or reversal) phase. For example, O'Leary, Becker, Evans, and Saudargas (1969) included several experimental phases in evaluating the effect of token reinforcement on classroom deportment. Separate phases included the effect of rules alone, educational structure, praise for appropriate and extinction of inappropriate behavior, and token reinforcement. After the token phase, a reversal of conditions was effected. The reversal condition (usually a return to baseline) is an essential ingredient in this design. Only such a reversal can demonstrate that behavior changes only when the experimental condition is in effect. This design is quite powerful and rules out several alternative explanations that may account for behavior change.

In spite of the usefulness and power of the design, it makes a major presupposition, namely that behavior changes made under various experimental conditions are reversible when baseline conditions are reinstated. The demonstration of a functional relationship between the presence of the experimental condition and performance requires that the changes made be transient and therefore reversible. However, the changes made in reinforcement programs might not be reversible when the experimental condition is withdrawn (*e.g.*, Surratt, Ulrich, and Hawkins, 1969). Indeed, we would hope that they are not always reversible, because this is tantamount to demonstrating only slight resistance to extinction.

If the effects of a reinforcement program are not reversible, the effect of the experimental condition(s) is not clear. For example, Hewett,

Taylor, and Artuso (1968) employed six classes in evaluating token reinforcement programs. Some classes received 17 weeks of baseline followed by 17 weeks of the program; other classes received 17 weeks of the program followed by 17 weeks of baseline (*i.e.*, no program). Two other classes were included in the design, to wit 34 weeks of the program and 34 weeks of baseline, respectively. For our purpose, it is important to note that the two groups that received the token program followed by baseline conditions failed to show a decline in target behaviors. Further, appropriate behaviors *increased* when the programs were removed. While speculations may be given to account for this (such as claiming that stimuli other than tokens in the environment have become reinforcers and maintain behavior), it is unclear that reinforcement was controlling any behavior in these two classes because no reversal occurred when the contingencies were removed. Addressing themselves to this problem, Bijou and associates (Bijou *et al.*, 1969) recommended using short experimental periods, which would facilitate obtaining a reversal of effects. This recommendation is useful when the goal is to determine short-term effects of experimental conditions. However, this design discussed (*i.e.*, ABAB) by itself, may be inadequate when one wishes to study non-transient effects. For example, if an investigator desires to implement a condition that leads to relatively permanent changes in behavior, this is difficult to demonstrate in this design. Hence, if an investigator adds to some experimental condition, a variable that should enhance resistance to extinction, this cannot be clearly demonstrated in this design. When behavior does not reverse, other explanations for the effect remain tenable. However, variables that may result in non-transient effects can be evaluated *after* a reversal phase has been successfully employed (Kazdin and Polster, 1973).

Recently, other within-group designs have been discussed that are not susceptible to the reversibility problem outlined above. These designs are particularly useful in situations where:

effecting a reversal would be undesirable because of exigencies of the situation; a reversal in responses would not be expected, as in training competence in academic skills; or where an experimental condition used is expected to inhibit a reversal of responses. The multiple-element baseline design (Sidman, 1960) or multiple-baseline design (Baer *et al.*, 1968) provides a valuable alternative to the ABAB design. In this design there is no reversal of conditions required to demonstrate the efficacy of the contingencies. Instead, data are collected across behaviors, across individuals, or across situations.

In the multiple-baseline design across behaviors, two or more behaviors are observed for the subject(s). After the behaviors have reached stable rates, the experimental condition is implemented for only one of the behaviors while baseline conditions are continued for the other(s). The behavior exposed to the experimental condition should change while the other behavior remains at baseline levels. When rates are stable for both behaviors, the second behavior is brought into the contingency. This procedure is continued until all behaviors for which multiple-baseline data were gathered are sequentially brought into the contingency. Ideally, each behavior changes only as it is included in the experimental contingency and not before. This is a powerful demonstration that the experimental condition exerts control over the behavior. The strength of the demonstration stems from the consideration that events occurring in time other than the experimental condition cannot plausibly account for the specific changes in behavior. This is demonstrated without a reversal of experimental conditions.

A major area of concern in using this design is that one must be reasonably assured beforehand that the target behaviors used are not interdependent or interrelated highly with each other. In such a situation, implementing a contingency for the performance of one behavior may be expected to alter the behavior(s) for which continued baseline data are collected. For example,

in a classroom situation it may not be the most appropriate design to gather multiple-baseline data across inappropriate motor behavior, inappropriate verbalizations, and inappropriate tasks as three separate target behaviors. Although these behaviors are used as distinct categories that can be reliably observed, they are also moderately intercorrelated in terms of frequency within individual children's repertoires (Kazdin, 1973c).² Change in one of these responses may result in other response changes. An even greater demonstration of this problem of response correlations is evident from the literature on generalized imitation. In several studies, it has been shown that reinforcement for some imitative behavior leads to a generalized set for imitative behaviors. A multiple-baseline design across behaviors might not be able to demonstrate that responses are not imitated until the contingency is applied to the specific imitative behavior. As soon as some imitative behavior is reinforced, other responses would change even though not reinforced (Baer, Peterson, and Sherman, 1967; Metz, 1965; Peterson and Whitehurst, 1971), unless they are topographically dissimilar responses (Garcia, Baer, and Firestone, 1971).

The problem of intercorrelations among responses has not been evident in the few multiple-baseline studies across behaviors (*e.g.*, McAllister, Stachowiak, Baer, and Conderman, 1969; Wolf *et al.*, 1968). However, it is a consideration an investigator should make when deciding on the type of design to best demonstrate the efficacy of experimental conditions (*cf.* Buell, Stoddard, Harris, and Baer, 1968; Pendergrass, 1972).

²This problem stems, in part, from the definition of an operant. As a response class, any operant may include responses or elements that are part of other operants. Skinner (1953, p. 94) noted this in stating that: "In reinforcing one operant we often produce a noticeable increase in the strength of another." and "The reinforcement of a response increases the probability of all responses containing the same elements." For a discussion of problems associated with defining operants, the reader is referred to Schick (1971).

In the multiple-baseline design across individuals, baseline data are gathered for at least one behavior across several persons. After behavior stabilizes across subjects, the experimental condition is invoked for one subject while baseline conditions are continued for the other subject(s). Again, as the experimental condition is extended to include separate individuals, the response frequency changes. This demonstration shows that behavior of the subject does not change until he is included in the experimental condition.

As with the previous multiple-baseline design, one major aspect may make this design problematic. If it is possible that the alteration of the behavior of one subject will influence the behavior of other subjects, the design loses power. In this situation, implementation of the experimental condition for the first subject may dramatically alter the behavior of another subject for whom baseline conditions are continued. For example, Broden, Bruce, Mitchell, Carter, and Hall (1970) showed that altering the attentive behaviors of one subject in the classroom through contingent reinforcement changed behavior of an adjacent peer. Further, data they collected indicated that when the attentive behaviors were changed in one subject, the source of distraction and possibly social reinforcement for the other subject decreased, so he improved because there was less opportunity and peer reinforcement for inattentiveness. Similarly, Kazdin (1973b) demonstrated that contingent reinforcement of attentive behavior for a target subject increased attentive behavior in an adjacent subject. However, it was also shown that reinforcement of inattentive behavior in the target subject reliably *increased* attentive behavior in the non-target subject. The discriminative stimulus value of social reinforcement appeared to account for the improvement in behavior of the adjacent subject. The point of these studies is to demonstrate that introducing contingencies for some individuals in a situation may be expected to alter the behaviors of other individuals under some circumstances, *viz.*, when the behavior of

one subject influences the behavior of an adjacent peer (Broden *et al.*, 1970) or when social reinforcement provides a discriminative stimulus for probable reinforcement for an adjacent peer (Kazdin, 1973*b*) or when there is a limit to the amount of reinforcement available (Sechrest, 1963). In such situations, the use of a multiple-baseline design across individuals would not effectively demonstrate the specific effects of the experimental condition on the target behavior. The implementation of the contingency for the first subject might change the behavior of others, even though the baseline conditions were continued for the other subjects.

In the multiple-baseline design across situations, data are collected for a target behavior for one or more subjects across different circumstances or situations. For example, in altering the promptness of individuals in an elementary school situation, one might collect data (number of students late and number of minutes late) across several situations (arrival to class in the morning, after recess, after lunch, after assemblies). After collecting baseline data in all situations, the experimental contingency is instituted to control the behavior in one situation. Baseline data are continued for behavior in all other situations until each is consecutively included into the contingency. As with previous multiple-baseline designs, this design is more effective when there is little or no correlation of behavior across these situations. If behavior change in one situation is expected to alter behavior in another situation, this design is a less powerful demonstration of the effects of the contingencies. For example, Hunt and Zimmerman (1969) demonstrated that contingent reinforcement for productivity in one time period increased productivity in a time period in which no reinforcement was delivered in a simulated sheltered workshop. There may be some question here as to the specific operation of the contingencies.

In the three basic types of multiple-baseline designs discussed, each has one potential weakness in powerfully demonstrating the effect of a

particular experimental condition, *viz.*, concomitant changes in the areas for which baseline data are collected. Whether this is the case in the particular instance the investigator decides to use a design must be determined primarily from experience. Concomitant changes that may occur as a result of implementing a contingency for one behavior (individual or situation) have to be determined empirically. In some instances, the investigator can rely on the well-documented experience of others. For example, if one were interested in evaluating classroom deportment in a multiple-baseline design across situations or time (such as morning and afternoon class periods), there already is consistent evidence showing that changes made in one of these time periods do not appreciably alter behavior in the other (Becker, Madsen, Arnold, and Thomas, 1967; Kuypers, Becker, and O'Leary, 1968; Meichenbaum, Bowers, and Ross, 1968; O'Leary *et al.*, 1969).

The within-subject designs outlined above can be employed in most situations. In spite of the reasons outlined early in the paper as to the advantages of these designs, one must remain cognizant of the possible interaction obtained between the experimental manipulation and the design employed in determining behavior. There is evidence bearing on this from several quarters in the experimental literature. For example, in experiments varying conditioned stimulus intensity in eyelid conditioning and signal intensity in reaction time experiments, the effects of variations of the stimuli depend upon whether they are evaluated within or between groups (Grice and Hunter, 1964). Similarly, the effects of different amounts of reinforcement in discrimination learning tasks differentially affect correct responses, again, depending on whether one group is exposed to different levels of reinforcement or separate groups are used (Lawson, 1957; Schrier, 1958). These experiments and others indicate that the effects obtained may be dependent upon the design of the experiment. There is a greater lesson to be learned from this than the simple recommendation that one

should not become overly dependent on one design in examining the effects of a particular operation. Since the major interest in operant work is determining how environmental manipulations functionally control behavior, it is important to be more analytic about an experimental design that, in part, dictates the results. What about the design influences behavior, and how can these influences be brought under control, minimized, or altered? When an experimental design is examined in this light, it becomes another experimental operation that exerts some functional relation to behavior. When the design does determine the result in some way, it is important to determine precisely how it accomplishes this and over what parametric levels of the experimental manipulation. Sidman (1960, pp. 334-340) recommended comparing the effects of experimental manipulations when scheduled separately or as combined with other manipulations. This is similar to recommending a closer scrutiny of the effects of our designs.

Evaluation of Results

Investigations of reinforcement programs in applied settings may introduce problems in evaluating the results. Initially, to discuss some of these problems, the distinction introduced by Campbell and Stanley (1963) on the validities of experimentation is useful. These authors refer to *internal validity* as the degree to which the results of an experiment are considered to be due to the experimental manipulation. *External validity* refers to the extent to which the findings obtained in a study may be extended or generalized to other groups and settings. Campbell and Stanley (1963) noted that the equivalent time-samples design (a version of the ABAB or reversal design) is quite strong with respect to the possible sources of threat to internal validity. In such designs, several rival hypotheses accounting for the results are ruled out. It is unlikely that events (outside of the experimental manipulation) that occur in time (history), growth or developmental processes within the subject (mat-

uration), systematic shifts in performance over time resulting from the unreliability of measurement (e.g., regression and changes in the measurement device), selective loss of subjects (mortality), repeated assessment (testing), and other factors account for the results. As for external validity, however, they list some factors that may delimit the generalizability of the results. Both these types of validity will be discussed in light of research in reinforcement programs in applied settings.

Considering internal validity, investigators must be reasonably assured that there are no factors in the design that can account for the results other than the intended manipulated operation. This statement seems so basic that it might not warrant protracted discussion. However, in several studies it is evident that there are factors that covary with implementation of the experimental operation. In some instances, the operation of these extraneous factors plausibly are causative of the changes attributed to the experimental operation. If these extraneous factor(s) could not account for the change entirely, they can interact with the experimental operation as a codeterminant of the results. Examples of extraneous factors that covary with experimental conditions are evident in several studies. One obvious factor that covaries with conditions is instructions that convey to subjects how they are supposed to perform. For example, Ayllon and Azrin (1965) instructed subjects that they could continue to work even though they would receive token reinforcement for not working (i.e., a "vacation with pay", p. 366). The rapid changes noted in performance were attributed to the reinforcement contingencies. However, different instructions preceding each experimental phase appeared to contribute, in part, to the abrupt effect of contingent reinforcement. Previous work has shown that although instructions may not be sufficient to sustain performance relative to contingent operant consequences, they are effective in initiating behavior change (Ayllon and Azrin, 1964; Hopkins, 1968; Packard, 1970). Em-

playing reinforcement contingencies alone results in behavior changes that are less dramatic than when accompanied by contingency instructions (Herman and Tramontana, 1971), although there are exceptions to this (Kazdin, 1973c).

The role of instructions is especially important to distinguish from that of reinforcement when the target behaviors are not apparent in the repertoire of the subjects. In this situation, the effect of reinforcement may be distinguished from initial training or practice in the reinforced responses. These are usually confounded in reinforcement studies. In one exception (Suchotliff, Greaves, Stecker, and Berke, 1970), psychiatric patients were exposed to training for grooming skills (where individuals were exposed to classes instructing them in the execution of the skills). Subsequently, a token reinforcement system was instituted to reinforce these behaviors. The results indicated that grooming behaviors did not increase during token reinforcement relative to what they were during training. Token reinforcement maintained the behaviors developed during the formal instruction period. However, the increments in performance over baseline were due to training and instruction. In several studies, the effect of instructions, in this sense, cannot be separated from the reinforcement contingencies themselves.

An additional factor when evaluating reinforcement is the possible experimental operations that are included in the effect of the contingent application of an incentive *per se*. The introduction of reinforcement for appropriate behavior includes several factors that lead to behavior change. For example, aside from an incentive function, reinforcement also serves an informational function (Bandura, 1971). Effects obtained with reinforcement may be unjustifiably attributed to the incentive function of consequating events, rather than the informational value of these events. Yet, providing feedback alone to a subject about the adequacy of his performance may affect performance dramati-

cally. While this may demonstrate that feedback serves as reinforcement for behavior (Mathis, Cotton, and Sechrest, 1971, p. 77), this does not help when investigators wish to make conclusions about the effect of, say, token reinforcement in changing behavior. The introduction of token reinforcement includes both the effect of incentives and informative feedback. These may be reduced to at least two distinct experimental operations. For example, Zimmerman *et al.* (1969) evaluated independently the effects of feedback and token reinforcement on productivity of clients in a sheltered workshop. Before the token reinforcement phase, subjects were told that they could practise (by working well) so that they would know how to earn tokens. During this phase, subjects were told how many tokens they would have earned, if tokens were given out. Subsequently, a token reinforcement phase was instituted. Feedback alone increased production, and token reinforcement resulted in further increases. Although token reinforcement was more effective, few studies have used as a base for comparison the feedback that is implicit in token reinforcement. Hence, conclusions about the effect of token reinforcement *per se* (particularly those made about the magnitude of change) often cannot be made on the basis of the experimental manipulation.

Instructions to agents who are administering the reinforcement program may also covary with the implementation and removal of the reinforcement contingencies. For example, teachers may be told or be led to believe that certain experimental phases will result in dramatic changes in behaviors. This may change teacher behavior, which influences the target behaviors. Studies have not carefully examined this.

In considering external validity, several issues in reinforcement programs require mentioning. The question of the extent to which findings hold for other subjects and other settings than those that were included in an experiment are included in the issues discussed here.

First, there is a possibility of *multiple-treatment interference* (Campbell and Stanley,

1963), which may delimit generalization of the results. Whenever multiple phases are applied to a group, the conclusions derived from a later treatment may depend on previous phases because the effects of each are not erasable. For example, the effects of contingent reinforcement may be evaluated in a design in which it is preceded by noncontingent reinforcement, or punishment (*e.g.*, verbal reprimand by teachers). The results can only be generalized to include other individuals exposed to a similar sequence of events. Also, the conclusions may be generalized only to conditions where the reinforcement condition is introduced repeatedly, interspersed with other conditions, and not to situations in which the reinforcement is continually present or introduced only once. For example, in two of their experiments, O'Leary and Becker (O'Leary and Becker, 1967; O'Leary *et al.*, 1969) implemented token reinforcement programs in classroom situations. In one study, instructions, praising appropriate and ignoring inappropriate behaviors along with token reinforcement, were introduced simultaneously; in the other study, these procedures were introduced sequentially in a cumulative fashion. In comparing these studies, it is evident that the simultaneous introduction of the conditions led to greater change than the sequential introduction of component parts. Thus, in generalizing the effects of token reinforcement procedures, one must be careful to specify the manner in which the component parts of the procedures (informative feedback, instructions, approval, ignoring) are introduced. The manner in which the program is introduced, as dictated by the experimental design, may have consequences as to the conclusions derived and their generalizability. In laboratory studies of punishment, for example, the manner in which the aversive event is introduced, high intensity initially or gradual approximation of high intensity, dictates the efficacy of the procedure (Azrin and Holz, 1966, p. 393).

The problem of multiple-treatment interference arises also in situations in which a base-

line (or reversal) is interspersed with different experimental conditions. If different versions of a reinforcement condition are presented and preceded by a baseline condition, it is difficult to separate the effects of the particular order of conditions from the effect of the conditions. Again, any conclusions reached about a particular condition are restricted to those individuals who are exposed to the particular sequence of conditions. When multiple reversals are included in the design, it is often assumed that if behavior returns to baseline levels, that any new condition may be compared to any other condition preceded by performance because of the equivalence of performance during reversal phases. There are several problems in comparing treatments within subjects related to the sequence effects. First, equivalent performance during reversals only demonstrates an equivalence in response rates at the time the reversal phase is in effect. However, reinstating the target behavior may become easier after repeated exposure to experimental conditions. It has already been demonstrated that repeated exposure to a series of alternating conditioning and extinction trials results in a gradual decline in responding over the extinction trials (Perkins and Cacioppo, 1950). A similar phenomenon, no doubt, exists in the repetitive exposure to the reconditioning phases themselves.

A related factor that affects the external validity of experiments in applied settings has been called *reactive effect of experimental arrangements* (Campbell and Stanley, 1963). This refers to the fact that the particular experimental arrangement may preclude generalization of treatment effects across time, situations, and individuals. As this relates to reinforcement programs, several aspects delimit generalization of the effects of the program. The definition of generalization varies here depending upon the particular problem considered. Initially, the development of reinforcement procedures in applied settings is usually accomplished by having observers record target behaviors of the clients or subjects. Although the influence of observers

is assumed not to affect the differential performance of, for example, school students, the presence of observers at all may constitute a reactive arrangement that delimits generalization of the findings to situations where observers are present. In classroom settings, both teacher and students are usually aware of the presence of the observers and may perform differently in situations where the observers are absent. There is only sparse evidence bearing on this point. Surratt *et al.* (1969) obtained data indicating that the presence of an observer in the classroom is a sufficient discriminative stimulus for performance of the target response. At the end of a reinforcement program designed to increase "time working" in students, two postchecks were conducted to determine to what extent the behaviors were maintained. The data from the first check were gathered by TV camera. Data for the second postcheck were gathered by an observer whose presence was previously associated with the experimental condition. "Time working" was consistently higher (in all four subjects) in the data collected by the observer. This suggests that there might be reactive effects of using observers in applied settings, particularly if the experimental conditions become associated with their presence. Subsequent introduction of an observer (S^D) may occasion the target behaviors, and the findings cannot be generalized beyond situations in which the observer is present.

A related factor not to be neglected is the effect of the investigator on the behavior of agents delivering reinforcement. The presence of the experimenter in the situation may occasion the desired response. Again, considering the classroom situation, the presence of the experimenter may serve as a reminder to the teacher to deliver reinforcement or to increase her rate of social approval. (When it is evident that observers are recording teacher behaviors, observers may also control these responses.) The major point is that the results obtained may be limited to those situations in which the experimenter is present (either in the situation itself

or available for consultation). Although the presence of the experimenter has been sparsely evaluated in reinforcement programs directly, there is evidence showing that the experimenter's presence may do this in other contexts. Peterson and Whitehurst (1971) showed that systematic variation of experimenter's presence exerts functional control over imitative behavior in children.

The Use of Control Groups

In discussions of work in the functional analysis of behavior in applied settings, the matter of control groups arises infrequently. There is great justification for ignoring the use of comparative groups in most instances. Treatises on experimental design (*e.g.*, Underwood, 1957) usually recognize certain instances in which no control group is required to evaluate unambiguously the effect of the experimental manipulation. The design that fits this situation is an extended time-series design in which data are available for the subject(s) over a long period of time, or when data are available for a relatively short period, but behavior changes with the presentation and removal of the experimental variable (*i.e.*, the ABAB design discussed earlier).

Other reasons have led investigators to eschew comparison groups stemming from philosophical or presuppositional considerations rather than simply convenience or design. The use of comparison groups usually implies a statistical evaluation of the data in terms of measures of central tendency and variability. Several assumptions are required for use of various statistical techniques and risks attendant on their violation must be considered. Further, in the experimental analysis, the goal transcends obtaining mean differences between groups with exposure to an experimental operation and one that is not. Achieving functional control over behavior makes the investigator concerned with determining effective variables that will alter the individual's behavior. These variables necessarily entail within-subject manipulations. Also, ex-

amining group means does not really demonstrate that the experimenter achieved control over behavior in individual cases. There is the subject generality problem (Sidman, 1960) or representativeness of the findings. Comparisons of treated groups with untreated groups obscure a closer examination of the effect of treatment on those individuals in the experimental groups. Nevertheless, there are instances, particularly in recent years, that merit utilization of experimental groups in between- as well as within-subject comparisons. The salient instances will be presented.

Whereas the initial interest in operant work in applied settings was focused almost entirely upon examining effective experimental operations that functionally related to behavior, this interest has broadened (Staats, 1970). There is greater use of finding effective treatments that cannot easily be evaluated in single-subject or single-group designs. For example, Staats and his associates began work with reinforcement procedures on training reading skills by a careful scrutiny of the performance of individual subjects (*e.g.*, Staats and Butterfield, 1965). This led to refinement and extensions of procedures which, though modified, were evaluated (statistically) as treatments and compared with a control group (*e.g.*, Staats, Minke, and Butts, 1970). The extension of findings from an examination of single-group to between-group designs has been required in light of the greater aims of reinforcement procedures in applied settings. As a recent example of this, initial token reinforcement programs in psychiatric hospitals (*e.g.*, Ayllon and Azrin, 1965) focused on evaluating the program on within-hospital behavior rather than global measures such as discharge and readmission rates (Ayllon and Azrin, 1968, p. 27). However, as the efficacy of the procedures evolved, it became important to determine how effective reinforcement procedures were relative to traditional techniques (Birky, Chambliss, and Wasden, 1971; Marks, Sanoda, and Schalock, 1968; Hartlage, 1970) and how well they fared when compared to untreated

groups in terms of follow-up success (Stayer and Jones, *unpublished*). Such comparisons, of course, imply the use of control groups.

Another situation that may require the use of comparison groups is related to the above. When different levels of variables are evaluated, the within-group design may restrict the external validity of the results. For example, if the effect of praise is evaluated and one wishes to compare that with the effect of praise and token reinforcement combined, the goal may not be achieved by merely introducing praise in one phase and adding token reinforcement in a subsequent experimental phase. The question of interest may be the effect of praise compared to that of praise-plus-token reinforcement when both are introduced initially to a group not previously exposed to any experimental phase. This requires separate groups.

In advocating the use of control groups, it is important to recognize the limitations imposed by doing research in applied settings. Even if we might envision situations in which a comparison group would provide desirable information, there are usually restrictions as to the information it can provide. With relatively rare exceptions in the literature (*e.g.*, Herman and Tramontana, 1971), subjects cannot be matched and assigned randomly to classes, hospital wards, institutional settings, or classrooms in which the procedures will be evaluated. In such instances it is desirable to select groups that will best control for the factors one is interested in controlling. For example, in a psychiatric setting, it is desirable to control for new staff, new ward facilities, and diagnostic group in evaluating the program. Some investigations in psychiatric settings have selected patients for the therapeutic program and placed them on a special ward (Heap, Boblitt, Moore, and Hord, 1970). This makes evaluation of any program ambiguous because the effect of ward change, in and of itself, may lead to behavior change (De Vries, 1968; Higgs, 1970). An excellent example of avoiding some of these problems can be found in the study by Schaefer and Martin

(1966). The patients on a ward were randomly assigned to either receiving contingent or non-contingent token reinforcement. Not only did controls receive tokens and live on the same ward as the experimental subjects, they also received contingent praise for desirable behaviors. This type of control group effectively precludes a number of factors that may account for differences between groups.

The use of controls may also be crucial when the experimenter wishes to determine whether relatively permanent effects of exposure to treatment conditions are obtained and the magnitude of these effects. For example, Wolf *et al.* (1968) evaluated the effects of token reinforcement in altering academic skills in a remedial classroom. Although the results were dramatic in within-group comparisons, the effectiveness of the program was enhanced by comparisons of students who received the program with those who did not. Over the period in which the program was evaluated (1 yr), both groups made significant gains on a standard achievement test and grades. However, the experimental group made substantial gains over and above those of controls. Clearly, a control group helped determine the magnitude of the gains made as a result of the program.

Responses Assessed in Reinforcement Programs

The selection of responses to evaluate the efficacy of procedures is usually dictated by the purpose of the study and the goals of the treatment or training institution in which the program is conducted. The use of response frequency has been employed as the most useful measure of these responses, and its use has a number of features to recommend it (see Bijou *et al.*, 1968, 1969; Ferster, 1953; Honig, 1966).

The focus on observable behaviors is perhaps one of the major advantages that accrue to behavior modification procedures in general. This assessment procedure differs markedly from traditional approaches where inferential leaps

may be made in diagnosing, treating, and assessing dynamics, dispositions, or traits (Mischel, 1968; Stuart, 1970).

In spite of the advantages of this assessment approach, the observations have been restricted to the single target behavior of initial focus. While changes in target behaviors are the *raison d'être* for undertaking treatment or training programs, concomitant changes may take place as well. If so, these should be assessed. It is one thing to assess and evaluate changes in a target behavior, but quite another to insist on excluding non-target measures. It may be that investigators are short-changing themselves in evaluating the programs. Recently, other areas of behavior modification have attempted to assess behaviors that might change as a result of treatment but were not of direct therapeutic focus (*e.g.*, Bandura, Blanchard, and Ritter, 1969; Kazdin, 1973a; Paul, 1967). To reiterate, the use of non-target behavioral measures in reinforcement programs is to be encouraged. Yet, these measures are not to be made at the expense of the primary data on target performance.

There are several potential advantages in using measures of non-target behaviors as well as the usual target response measures. One initial advantage is that such assessment would permit the possibility of determining response generalization. If certain response frequencies are increased or decreased, it would be expected that other related operants would be influenced. It would be a desirable addition to determine generalization of beneficial response changes by looking at behaviors related to the target response. In addition, changes in the frequency of responses might also correlate with topographical alterations. For example, a reduction of inappropriate responses may result in the concomitant reduction of the severity of the responses (Burchard and Tyler, 1965; Hawkins, Peterson, Schweid, and Bijou, 1966).

As the application of reinforcement programs proliferates further, there will be increased concern with concomitant effects of such programs, aside from their obvious virtue of reliably alter-

ing the target behavior(s). Examples of this can be seen in a few studies. Mulligan, Kaplan, and Reppucci (1971) evaluated changes in "cognitive" variables made as a result of a token program for appropriate classroom behaviors and completion of tasks. The elementary school subjects in the program showed increases in IQ scores, arithmetic achievement scores, and a slight decrease in anxiety. The conclusions reached were that cognitive as well as behavioral improvement may occur in a token reinforcement program. (This particular study has methodological problems such as possibility of regression artifacts contributing to the results.) This suggests that there were important gains made outside of the obvious target response changes.

In using non-target as well as target measures to assess changes made in reinforcement programs, it is important to maintain emphasis on the target behavioral measures. It is this measure that can determine whether the program has functional control over behavior. For example, Gripp and Magaro (1971) compared schizophrenic patients on one ward exposed to token reinforcement with control wards not exposed to this treatment. Ratings were made on several indirect measures of psychotic behavior, ward atmosphere, and social behaviors. Conclusions were made on a number of dimensions relevant to psychopathology. The comparisons were determined from preprogram and postprogram ratings. The experimental ward showed greater improvement than controls in number of dimensions of gains and in degree of gains. However, because the investigation did not focus on direct behavioral targets on which the program could be evaluated, and hence could not demonstrate functional control, the conclusions are ambiguous. The effect of the program may have been due to collateral changes in staff behaviors, and a host of other factors. The use of indirect or multiple measures of treatment efficacy are only relevant when it has been demonstrated that the contingencies themselves are responsible for the direct changes in behavior. When

the program has demonstrated control over behavior, it would be interesting to determine non-target correlates that may change as well.

SUMMARY

In summing up, there are several advantages in using various within-subject designs. Generally, these designs are quite powerful in demonstrating the effect of a particular experimental operation. However, there are potential weaknesses of the designs currently employed. The choice of design may be influenced by the expectation or desirability of a reversal in behavior, if the experimental condition is withdrawn. In designs without a reversal of conditions (multiple-baseline designs), other problems may arise, such as interdependence of performance (across behaviors, individuals, or situations). Although these may prove to be infrequent in future research, their occurrence in a given instance may be fatal in evaluating the results.

In interpreting results of investigations in this area, it is important to be cognizant of potential influences that may covary with the presentation and withdrawal of experimental operations. Also, various elements of the experiment may delimit generalization of the results. Salient influences relevant to internal and external validity were discussed.

The use of control groups was advocated in examining certain effects of reinforcement programs. Although the functional analysis is the major aspect of the design, the proliferation of experiments in this area has led to questions that can, in many instances, be adequately answered only by comparisons between or across groups. As a final point, the use of multiple-response measures was encouraged. Changes in target measures are the major point of undertaking reinforcement programs in applied settings. However, concomitant response changes may take place along with the target responses. If so, these should be documented.

REFERENCES

- Ayllon, T. and Azrin, N. H. Reinforcement and instructions with mental patients. *Journal of the Experimental Analysis of Behavior*, 1964, 7, 327-331.
- Ayllon, T. and Azrin, N. H. The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior*, 1965, 8, 357-383.
- Ayllon, T. and Azrin, N. H. *The token economy: a motivational system for therapy and rehabilitation*. New York: Appleton-Century-Crofts, 1968.
- Azrin, N. H. and Holz, W. C. Punishment. In W. K. Honig (Ed.), *Operant behavior: areas of research and application*. New York: Appleton-Century-Crofts, 1966. Pp. 380-447.
- Baer, D. M. Some remedial uses of the reinforcement contingency. In J. Shlien (Ed.), *Research in psychotherapy, Volume III*. Washington, D.C.: American Psychological Association, 1968. Pp. 3-20.
- Baer, D. M. Behavior modification: you shouldn't. In E. A. Ramp and B. L. Hopkins (Eds.), *A new direction for education: behavior analysis*, 1971. Lawrence, Kansas: University of Kansas Support and Development Center for Follow Through, 1971. Pp. 358-369.
- Baer, D. M., Peterson, R. F., and Sherman, J. The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior*, 1967, 10, 405-416.
- Baer, D. M., Wolf, M. M., and Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, 1, 91-97.
- Bandura, A. *Principles of behavior modification*. New York: Holt, Rinehart, & Winston, 1969.
- Bandura, A. Vicarious- and self-reinforcement processes. In R. Glaser (Ed.), *The nature of reinforcement*. New York: Academic Press, 1971. Pp. 228-278.
- Bandura, A., Blanchard, E. B., and Ritter, B. Relative efficacy of desensitization and modeling approaches for inducing behavioral, affective, and attitudinal changes. *Journal of Personality and Social Psychology*, 1969, 13, 173-199.
- Becker, W. C., Madsen, C. H., Arnold, C. R., and Thomas, D. R. The contingent use of teacher attention and praising in reducing classroom behavior problems. *Journal of Special Education*, 1967, 1, 287-307.
- Bijou, S. W., Peterson, R. F., and Ault, M. H. A method of integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis*, 1968, 1, 175-191.
- Bijou, S. W., Peterson, R. F., Harris, F. R., Allen, K. E., and Johnston, M. S. Methodology for experimental studies of young children in natural settings. *Psychological Record*, 1969, 19, 177-210.
- Birky, H. J., Chambliss, J. E., and Wasden, R. A comparison of residents discharged from a token economy and two traditional psychiatric programs. *Behavior Therapy*, 1971, 2, 46-51.
- Brodin, M., Bruce, M., Mitchell, M., Carter, V., and Hall, R. V. Effects of teacher attention on attending behavior of two boys at adjacent desks. *Journal of Applied Behavior Analysis*, 1970, 3, 199-203.
- Buell, J., Stoddard, P., Harris, F., and Baer, D. M. Collateral social development accompanying reinforcement of outdoor play in a preschool child. *Journal of Applied Behavior Analysis*, 1968, 1, 167-173.
- Burchard, J. D. Systematic socialization: A programmed environment for the habilitation of antisocial retardates. *Psychological Record*, 1967, 17, 461-476.
- Burchard, J. D. and Tyler, V. O. The modification of delinquent behaviour through operant conditioning. *Behaviour Research and Therapy*, 1965, 2, 245-250.
- Campbell, D. T. and Stanley, J. C. Experimental and quasi-experimental designs for research and teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1963. Pp. 171-246.
- DeVries, D. L. Effects of environmental change and of participation on the behavior of mental patients. *Journal of Consulting and Clinical Psychology*, 1968, 32, 532-536.
- Ferster, C. B. The use of the free operant in the analysis of behavior. *Psychological Bulletin*, 1953, 50, 264-274.
- Garcia, E., Baer, D. M., and Firestone, I. The development of generalized imitation within topographically determined boundaries. *Journal of Applied Behavior Analysis*, 1971, 4, 101-112.
- Grice, C. R. and Hunter, J. J. Stimulus intensity effects depend upon the type of experimental design. *Psychological Review*, 1964, 71, 247-256.
- Gripp, R. F. and Magaro, P. A. A token economy program evaluation with untreated control ward comparisons. *Behaviour Research and Therapy*, 1971, 9, 137-149.
- Hartlage, L. C. Subprofessional therapists' use of reinforcement versus traditional psychotherapeutic techniques with schizophrenics. *Journal of Consulting and Clinical Psychology*, 1970, 34, 181-183.
- Hawkins, R. P., Peterson, R. F., Schweid, E., and Bijou, S. W. Behavior therapy in the home: Amelioration of problem parent-child relations with the parent in a therapeutic role. *Journal of Experimental Child Psychology*, 1966, 4, 99-107.
- Heap, R. F., Boblitt, W. E., Moore, C. H., and Hord,

- J. E. Behavior-milieu therapy with chronic neuropsychiatric patients. *Journal of Abnormal Psychology*, 1970, **76**, 349-354.
- Herman, S. and Tramontana, J. Instructions and group *versus* individual reinforcement in modifying disruptive group behavior. *Journal of Applied Behavior Analysis*, 1971, **4**, 113-119.
- Hewett, F. M., Taylor, F. D., and Artuso, A. A. The Santa Monica Project: Evaluation of an engineered classroom design with emotionally disturbed children. *Exceptional Children*, 1969, **35**, 523-529.
- Higgs, W. J. Effects of gross environmental change upon behavior of schizophrenics: A cautionary note. *Journal of Abnormal Psychology*, 1970, **76**, 421-422.
- Honig, W. K. Introduction. *Operant behavior: areas of research and application*. New York: Appleton-Century-Crofts, 1966.
- Hopkins, B. L. Effects of candy and social reinforcement schedule learning on the modification and maintenance of smiling. *Journal of Applied Behavior Analysis*, 1968, **1**, 121-128.
- Hunt, J. G. and Zimmerman, J. Stimulating productivity in a simulated workshop setting. *American Journal of Mental Deficiency*, 1969, **74**, 43-49.
- Kazdin, A. E. The effect of response cost and aversive stimulation in suppressing punished and nonpunished speech disfluencies. *Behavior Therapy*, 1973, **4**, 73-82. (a)
- Kazdin, A. E. The effect of vicarious reinforcement on attentive behavior in the classroom. *Journal of Applied Behavior Analysis*, 1973, **6**, 71-78. (b)
- Kazdin, A. E. The role of instructions and reinforcement in behavior changes in token reinforcement programs. *Journal of Educational Psychology*, 1973, **64**, 63-71. (c)
- Kazdin, A. E. and Bootzin, R. R. The token economy: an evaluative review. *Journal of Applied Behavior Analysis*, 1972, **5**, 343-372.
- Kazdin, A. E. and Polster, R. Intermittent token reinforcement and response maintenance in extinction. *Behavior Therapy*, 1973, **4**, 386-391.
- Kuypers, D. S., Becker, W. C., and O'Leary, K. D. How to make a token system fail. *Exceptional Children*, 1968, **11**, 101-108.
- Lawson, R. Brightness discrimination performance and secondary reward strength as a function of primary reward amount. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 35-39.
- McAllister, L. W., Stachowiak, J. G., Baer, D. M., and Conderman, L. The application of operant conditioning techniques in a secondary school classroom. *Journal of Applied Behavior Analysis*, 1969, **2**, 277-285.
- 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.
- Marks, J., Sonoda, B., and Schalock, R. Reinforcement *vs.* relationship therapy for schizophrenics. *Journal of Abnormal Psychology*, 1968, **73**, 397-402.
- Mathis, B. C., Cotton, J. W., and Sechrest, L. *Psychological foundations of education*. New York: Academic Press, 1971.
- Meichenbaum, D. H., Bowers, K., and Ross, R. R. Modification of classroom behavior of institutionalized female adolescent offenders. *Behaviour Research and Therapy*, 1968, **6**, 343-353.
- Metz, J. R. Conditioning generalized imitation in autistic children. *Journal of Experimental Child Psychology*, 1965, **2**, 389-399.
- Mischel, W. *Personality and assessment*. New York: Wiley, 1968.
- Mulligan, W., Kaplan, R. D., and Reppucci, N. D. *Changes in cognitive variables among behavior problem elementary school boys treated in a token economy special classroom*. Unpublished paper presented at Association for the Advancement of Behavior Therapy, Washington, D.C., September 1971.
- O'Leary, K. D. and Becker, W. C. Behavior modification of an adjustment class: A token reinforcement program. *Exceptional Children*, 1967, **9**, 637-642.
- O'Leary, K. D., Becker, W. C., Evans, M. B., and Saudargas, R. A. A token reinforcement program in a public school: A replication and systematic analysis. *Journal of Applied Behavior Analysis*, 1969, **2**, 3-31.
- O'Leary, K. D. and Drabman, R. Token reinforcement programs in the classroom: A review. *Psychological Bulletin*, 1971, **75**, 379-398.
- Packard, R. G. The control of "classroom attention": a group contingency for complex behavior. *Journal of Applied Behavior Analysis*, 1970, **3**, 13-28.
- Paul, G. L. Insight versus desensitization in psychotherapy two years after termination. *Journal of Consulting Psychology*, 1967, **31**, 333-348.
- Pendergrass, V. E. Timeout from positive reinforcement following persistent, high-rate behavior in retardates. *Journal of Applied Behavior Analysis*, 1972, **5**, 85-91.
- Perkins, C. C. and Cacioppo, A. J. The effect of intermittent reinforcement on the change in extinction rate following successive reconditionings. *Journal of Experimental Psychology*, 1950, **40**, 794-801.
- Peterson, R. F. and Whitehurst, G. J. A variable influencing the performance of generalized imitative behaviors. *Journal of Applied Behavior Analysis*, 1971, **4**, 1-9.
- Phillips, E. L. Achievement Place: token reinforcement procedures in a home-style rehabilitation setting for "predelinquent" boys. *Journal of Applied Behavior Analysis*, 1968, **1**, 213-223.

- Risley, T. R. Behavior modification: An experimental-therapeutic endeavor. In L. A. Hamerlynck, P. O. Davidson, and L. E. Acker (Eds.), *Behavior modification and ideal mental health services*. Calgary, Alberta, Canada: University of Calgary Press, 1970. Pp. 103-127.
- Risley, T. R. and Baer, D. M. Operant behavior modification: The deliberate development of child behavior. In B. Calwell and H. Riccui (Eds.), *Review of child development research, Volume III: Social action*, 1973, in press.
- Risley, T. R. and Wolf, M. M. Strategies for analyzing behavioral change over time. In J. Nesselroade and H. Reese (Eds.), *Life-span developmental psychology: methodological issues*. New York: Academic Press, 1972.
- Schaefer, H. H. and Martin, P. L. Behavior therapy for "apathy" of hospitalized schizophrenics. *Psychological Reports*, 1966, 19, 1147-1158.
- Schick, K. Operants. *Journal of the Experimental Analysis of Behavior*, 1971, 15, 413-423.
- Schrier, A. M. Comparison of two methods of investigating the effects of amount of reward on performance. *Journal of Comparative and Physiological Psychology*, 1958, 51, 725-731.
- Sechrest, L. Implicit reinforcement of responses. *Journal of Educational Psychology*, 1963, 54, 197-201.
- Sidman, M. *Tactics of scientific research*. New York: Basic Books, 1960.
- Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.
- Skinner, B. F. *Contingencies of reinforcement: a theoretical analysis*. New York: Appleton-Century-Crofts, 1969.
- Staats, A. W. Reinforcer systems in the solution of human problems. In G. A. Fargo, C. Behrens, and P. Nolen (Eds.), *Behavior modification in the classroom*. Belmont, California: Wadsworth, 1970. Pp. 6-31.
- Staats, A. W. and Butterfield, W. H. Treatment of nonreading in a culturally deprived juvenile delinquent: An application of learning principles. *Child Development*, 1965, 4, 925-942.
- Staats, A. W., Minke, K. A., and Butts, P. A token-reinforcement remedial reading program administered by black therapy technicians to problem black children. *Behavior Therapy*, 1970, 1, 331-353.
- Stayer, S. J. and Jones, F. Ward 108: *Behavior modification and the delinquent soldier*. Unpublished paper presented at Behavioral Engineering Conference, Walter Reed General Hospital, 1969.
- Stuart, R. B. *Trick or treatment: how and when psychotherapy fails*. Champaign, Illinois: Research Press, 1970.
- Suchotliff, L., Greaves, S., Stecker, H., and Berke, R. Critical variables in the token economy. *Proceedings of the 78th Annual Convention of the American Psychological Association*, 1970, 5, 517-518.
- Surratt, P. R., Ulrich, R. E., and Hawkins, R. P. An elementary student as a behavioral engineer. *Journal of Applied Behavior Analysis*, 1969, 2, 85-92.
- Thoresen, C. E. *The intensive design: an intimate approach to counseling research*. Unpublished paper presented at meeting of American Educational Research Association, Chicago, April, 1972.
- Underwood, B. J. *Psychological research*. New York: Appleton-Century-Crofts, 1957.
- Wahler, R. G. Setting generality: some specific and general effects of child behavior therapy. *Journal of Applied Behavior Analysis*, 1969, 2, 239-246.
- Wolf, M. M., Giles, D. K., and Hall, R. V. Experiments with token reinforcement in a remedial classroom. *Behaviour Research and Therapy*, 1968, 6, 51-64.
- Wolf, M. M. and Risley, T. R. Reinforcement: applied research. In R. Glaser (Ed.), *The nature of reinforcement*. New York: Academic Press, 1971. Pp. 310-325.
- Zimmerman, J., Stuckey, T. E., Garlick, B. J., and Miller, M. Effects of token reinforcement on productivity in multiply handicapped clients in a sheltered workshop. *Rehabilitation Literature*, 1969, 30, 34-41.

Received 21 March 1972.

(Revision requested 27 December 1972.)

(Final acceptance 29 January 1973.)