DOCUMENT RESUME

ED 178 843 CG 013 957

AUTHOR Gouvier, William Drew
TITLE Assessment of Change Agent Behavior: Design Strategies and Limitations.
PUB DATE [79]
NOTE 31p.; Paper presented at the Annual Convention of the Association for Behavior Analysis: International Organization (5th, Dearborn, Michigan, June 16-19, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Behavior Change; *Change Agents; *Evaluation Methods; *Interaction Process Analysis; Intervention; *Performance Factors; Psychological Studies; Rating Scales; Reinforcers; *Research Design

ABSTRACT

Relationships between change agent performance and outcome success within the context of group behavioral programs are identified. The evaluation of either the ultimate effectiveness, or the specific change processes in a group program often requires assessing the behavior of the change agents, as well as that of the target population. Five design strategies for conducting the assessment of change agents are presented, and limitations on the usefulness of each are noted. It is suggested that a greater amount of useful information can be gathered at a minimum of additional response cost by nesting a simple interaction scoring system into the appropriate categories of a multiple outcome assessment device. (Author)

Conference noted on release form. C.K. Jaslow

Assessment of Change Agent Behavior: Design Strategies and Limitations

William Drew Gouvier

Memphis State University

Assessment of Change Agent Behavior: Design Strategies and Limitations

> It is not the benevolence of the controller, but the conditions under which he controls benevolently which must be examined. (Skinner, 1968)

The applied behavioral literature contains well over a hundred studies concerning the training and utilization of non-professional change agents as primary service deliverers, program aides, behavioral observers and in adjunct functions in clinical research and therapy (see Berkowitz & Graziano, 1972; Gardner, 1973; O'Dell, 1974). These reviewers have pointed out that complex issues related to the evaluation of group interventions are frequently overlooked or ignored.

In presenting the rational for any group intervention, three broad areas deserve attention. First, the intervention must be related to some desired behavior changes among members of the target population. It is poor program planning to intervene with a technique that has been demonstrated to be inappropriate for that given population. Therefore, considerable research has attempted to define the populations most amenable to change via specific interventions. Secondly, the techniques of intervention must be presented to, and apprehended by the change agents. Hence, much research has been generated to study

the role of various instructional formats, content areas, and procedural parameters of the training package. These issues are well addressed in the reviews cited above. Lastly, the change agents must implement these intervention techniques correctly in order to maximize the therapeutic influence upon the target population (Kazdin, 1973). This issue has received limited attention; says Gripp (Gripp & Magaro, 1974), "Research on staff variables influencing token economy program effectiveness has not been done, and assertions concerning them are speculative and anecdotal."

The present paper will present an evaluation of current knowledge related to this third issue of change agent performance. The ultimate success of any given program depends on much more than the program's behavioral principals (Atthowe & Krassner, 1968). Five documented factors responsible for the failure of many behavior change programs include the following: (Andrasik & McNamara, 1977)

1) Inadequate responsiveness of the target population to the proffered reinforcers
2) Underdeveloped or uncreative behavioral innovation
3) Poor program design
4) Sociopolitical obstacles to change
5) Breakdown of the organization and internal control within the program system itself.

These authors point out that "A common element underlying many of these failures is the inadequate participation of the staff in performing program related duties." Observational studies of change agent behavior have indicated that agents often spend relatively little of their work time interacting with the target population in either training (Harmatz, 1973) or recreational activities (Daily, Allen, Chinsky, & Veit, 1974), while much of their time may be spent in self-centered "leisure time" activity (Bensberg & Barnett, 1966). Herein lies the fundamental relevance of assessing change agent behavior; how can program administrators acquire and maintain control over the behavior of the change agents (Tharpe & Wetzel, 1969)? The answer to this query is an empirical process of three stages.

1) The influences that control the change agent's behavior need to be identified first. These influences include the stimulus control of behavior, the contingencies relating responses to outcomes, the training and motivation of the agents, and the administrative controls that are operative at that time.

2) To effect the desired improvements in the agent's behavior, certain of these influences are systematically manipulated so that their relative importance can be assessed.

3) The altered behavior of the change agent is reassessed to allow for the demonstration of functional control exerted by the aforementioned influences, and also to determine the

utility of that change, as measured by behavioral improvements in the target population members. Additionally, it is only when such assessments are conducted can comparisons between intervention programs be interpreted with any degree of confidence (Milby, 1975).

The body of this paper will outline representative attempts to assess change agent performance through the aforementioned stages. Although the conceptual distinction has been made between the identification of variables that influence performance, and the assessment of behavior changes following the manipulation of these variables, one must bear in mind that the tasks of assessment and intervention are intertwined in practice (Andrasik & McNamara, 1977). Hence, the studies reviewed here will simply be presented in an ascending order of complexity, and only those studies which report reliability coefficients will be reviewed.

Designs

Broadly conceived, five strategies have been employed for the assessment of agent performance. These strategies range from indirect assessment of performance outside the job situation to complex records of change agent-target (e.g., staff-patient or teacher-child) interactions in vivo. Each area will be reviewed separately below.

Indirect Assessment

The use of objective and standardized behavioral tasks to assess the effectiveness of training programs has been reported by Gardner (1972) for staff behavior on a psychiatric ward, and replicated by Nay (1975) within the context of parent training. The factorial design of both of these studies contrasts two modes of training (role playing versus lecture) with two methods for assessing trainee competence (paper and pencil test versus ratings of role played behavior modification skills). These studies have demonstrated that roleplaying is a more effective method for imparting skill competence to the trainees, while a lecture format better facilitates performance on the verbally loaded paper and pencil test.

This information alone is interesting, but of little value unless the package's content is utilized in the change agents' routine interactions with members of the target population, and that the application of the new techniques is not limited solely to the training situation. The need to program generalization into the training package has been widely addressed (Horton, 1975; Kazdin, 1977; Miller & Sloane, 1976). Although Gardner (1972) reports moderately high correlations between ratings of roleplayed and in vivo performance ($r = .87$), there are no suggestions forwarded as to how this correlation could be improved, or what factors produced that correlation in the

first place. Therefore, it would be inappropriate to assume that correlations of this order represent typical examples of the relationship between behavior in vivo and that measured in an artificially constructed assessment situation. Other studies (Horton, 1975; Miller & Sloane, 1976) indicate that lesser degrees of generalization are in fact far more commonplace.

The additional complexity and redundancy invoked by this need for a separate assessment of generalizability is a serious flaw of indirect procedures. This shortcoming can be avoided only if behavior is measured in the performance situation as well. Since in vivo performance is the most valid measure of performance competencies (Bandura, 1977), indirect assessment devices might optimally be relegated to the role of screening out change agents not yet prepared for direct interaction with the target population, so that these agents can receive remedial training. The remainder of the studies reviewed conduct assessment of agent performance in vivo, either by assessing changes in the behavior of target population members (and inferring changes in staff behavior to be responsible for the observed changes), or by assessing the agents' behavior directly.

Single Target Assessments

Early studies utilizing nonprofessional change agents often targeted a single measure of patient change as the sole index of the effectiveness of training and intervention programs.

This procedure is insensitive to specific changes in the change agents' behavior which could be instrumental in producing the desired changes among the target population; hence, it is a procedure of little practical usefulness in assessing staff performance. The report of Bushell, Wrobel and Michaelis (1968) is typical of this approach. All student behavior was dichotomized as either studying or not studying. Improvements were noted in the proportion of time spent studying after the introduction of group contingencies to that end. These authors concluded that the contingencies were responsible for the observed changes, and that the teacher correctly applied the contingencies. Without appropriate design and controls, however, it is impossible to rule out alternate explanations based on demand characteristics or altered expectations of the teacher.

Avoiding some of these interpretational limitations, Broden, Bruce, Mitchell, Carter and Hall (1970) report a multiple baseline plus reversal design in their intervention upon the disruptive behavior of two students. Child behavior was dichotomized as either studying or not. Contingent attention for study was applied to one student, and then the other. Only the targeted child's behavior changed during this phase; return to baseline and reintervention phases produced the expected results. Even though this design allows stronger assertions about the control of student behavior than the Bushell et al.

(1968) report, it remains conceivable that some other change in teacher behavior was responsible for the observed changes in the target. Shortcomings of the reliance upon behavior changes in the target population for assessment of change agent performance have been addressed by Atthowe and Krassner (1968) and more recently by Gripp and Margaro (1974). Specifically, measures of patient change are insensitive to the particular changes in change agent performance, and hence are of limited utility in identifying the factors that control and maintain optimal staff performance, as well as being limited in their ability to identify particular change agent behaviors that are responsible for the desired changes among the target population members.

These problems are avoided in part by assessing the change agent behavior directly. Manelker, Brigham and Burnel (1970) measured changes in the frequency of "verbal statements directed to the child" as a function of teachers' implementing token reinforcement programs to reward in-class writing. Although these authors used a dubious technique of determining the reliability of their measurements (the behavioral observer rescored sessions from a video recording, after scoring them the first time), they report that the act of distributing contingent tokens to the children altered the behavioral environment in such a way as to strengthen those contingencies which

maintain teachers' speech. Similar increases in teacher verbalization were not noted with noncontingent token delivery or were instructions to attend more closely to the children.

Other authors have assessed the generalization of the change agents' application of specific techniques for which they had been trained. Utilization of techniques can be restricted to the environment in which the training occured (Miller & Sloane, 1976) or to the specific subject matter of the training (Horton, 1975). No significant generalization of training was noted in either of these investigations, although Horton reports that providing change agents with performance feedback does improve performance within the defined subject area.

These studies assessing a single behavior of the change agents or of the target population members have served to identify some of the variables that control change agent behavior. Reminders (Fielding, Erickson, & Beltin, 1977), performance feedback (Horton, 1975), rules for intervention (Bushell et al., 1968) and programmed generalization (Miller & Sloane, 1976) are all strongly implicated as factors that contribute to change agent performance. Despite refinements in the design of studies utilizing single target assessments, it is still not possible to partial out the effects of specific change agent activities from the influence of training program effectiveness. Such

effects can only be identified and studied when both target population and change agent behavior is assessed concurrently. Additionally, the limited focus of the aforementioned designs makes unlikely the possibility of identifying non-targeted covariate effects associated with intervention implementation (Wahler, 1968).

Permanent Product Assessment

In cases where there is a relationship between behavior changes and records of change agent activity, measures of these records can serve as indices of change agent functioning. It is not any behavior that is counted or rated, it is a measure of products resulting from appropriate and inappropriate change agent activity that is assessed. This type of assessment is useful both for identifying influences that control change agent behavior, as well as providing objective content for feedback to the change agents as a means of altering said behavior.

Barnard, Christophersen and Wolf (1974) report an investigation designed to correct problems in maintaining acceptable job performance among paraprofessional tutors in a reading program. Each day, one of five tutors was randomly selected to receive feedback from the program supervisor regarding the completeness of the tutoring, the correctness of the workbook scoring, and their promptness in reporting to work, using number of workbook pages completed, the scoring totals, and timeclock

cards as the source data. They report that this feedback procedure differentially influenced the three dependent measures, with completeness reaching asymtote for all tutors, significant but smaller improvements in accuracy, and no improvement in promptness. They conclude that completeness of performance is closely related to the extent that change agents are supervised. Furthermore, the use of the random feedback procedure appears to have prevented reactive effects of reliability checks on performance (Kazdin, 1977), because no tutor could know when evaluation would occur.

Andrasik and McNamara, (1977) sought likewise to correct poor performances in a "rehabilitation training facility." The target of this intervention was to eliminate instances in which token awards were made at times or in quantities other than those specified in each patient's contract. The effects and control exerted by several different variables were empirically demonstrated in a multiple baseline-reversal design. Merely by changing the format of the token "bankbooks" so that staff became accountable for their awards, the differential effects of positive feedback, punishment contingencies and recording format changes could be assessed precisely.

These authors assessed the possibility that a program employing aversive constraints over staff behavior might produce adverse deterioration in other job related activities.

Using measures from a staff attitude survey conducted by outside researchers, as well as frequencies of "abused" sick leave, they concluded that potentially adverse side effects had not been produced; and that policy changes, equipment design and information feedback were all partial determinants of staff performance.

An interesting demonstration of the multitude of ways simple records of permanent products can be used is provided by Coleman and Buren (1969). These authors substituted records of each patient's daily token exchanges for what were described as "negatively oriented nursing notes." Relief shift staff used these token account sheets to adjudge each patient's daily performance. By grouping these records according to various independent variables (i.e., treatment group, length of stay) useful feedback information could be provided to the staff regarding individual, group or ward operation. Longer term assessments of these operations could be obtained by graphing weekly or monthly expenses and earnings in cumulative blocks. This provides an index of ward operation likened to the way the Dow Jones Industrial Average represents fluctuations in the stock market.

Although such graphs do not provide an experimental demonstration of functional control of behavior, they do provide an effective use for feedback available from data review

techniques that can be used to improve and evaluate specific aspects of staff performance. Perhaps a more serious limitation inherent in the use of permanent products as a solitary assessment device is that it is a relatively blind assessment strategy. Consider a hypothetical training program to increase the frequency of diaper checks and changes on a ward composed of incontinent adults. A permanent product assessment might yield a significant increase in the number of soiled diapers washed daily, but it is quite possible that this improvement is accompanied by unexpected deleterious effects on the quality of the patient-staff interactions. Measurement insensitivity of this sort is a problem that plagues all of the assessment strategies reviewed so far. Ways to overcome this limitation include the use of additional assessment techniques or targets for two-fold purposes. First, the assessment data can be cross validated (in our previous example, the change agents could have soiled a clean diaper with each dirty one changed, and gotten credit of two for one.) Secondly, the influence of "side effects" (Wahler, 1974) can be identified and further investigated.

Multiple Measure Assessments

Andrasik and McNamara (1977) utilized outside personnel to assess potential changes in attitudes or job performance following their program implementation. It is becoming more

common for researchers to design their assessment strategies for the recording of multiple measures. As with single target assessments, multiple measure assessments of change agent performance can measure changes in the behavior of target population members, the behavior of the change agents, or combinations of both techniques can be employed. Both behavioral and permanent product records can be employed. Because of the complexities involved in designing and using a multiple measure procedure (Taplan and Reid, 1973), this technique has seen little use for the assessment of change agent behavior, despite the increase in useful information such procedures provide.

Hall, Panyan, Rabon and Broden (1968) report an early study in which both teacher attending and student studying behavior were dichotomized and measured. Following the training of the teacher to apply contingent attention to improve classroom studying, their measurements revealed concurrent changes in both the teacher's attending behavior and the class' studying throughout two baseline and two intervention phases. This concordance of behavior changes helps to demonstrate the functional interrelatedness of attention and studying, and the records of teacher attention provided the content of feedback provided to the teacher should performance fall below criterion levels. Since no other data were collected, however, one cannot

ascertain whether any concomitant changes accompanied or were responsible for the intervention's success.

Jones and Eimers (1975) sought to assess the efficacy of their performance oriented teacher training program by measuring three categories of disruptive child behavior, and one child produced permanent product. Following significant changes in these measures, these authors concluded that their package was effective, and that the teachers satisfactorily mastered and utilized its components. Again, it is not possible to assess the contribution of demand and social influence characteristics to this observed change, nor is it possible to identify concurrent non targeted changes in teacher behavior which may independently contribute to the obtained outcome. This is not a wholly damning criticism; for purposes of clinical utility it is often sufficient to demonstrate the effectiveness of any given procedure (outcome research). To understand the components of a procedure which are responsible for producing that change (process research) requires that the process of assessment be a more precisely empirical, and less speculative endeavour (Craighead, Kazdin & Mahoney, 1976).

A sophisticated example of the use of multiple measures in the assessment and intervention of change agent performance is reported by Iwata, Bailey, Brown, Foshee and Alpern (1976). These authors sought to improve patient dental care by using

access to preferred activities as reinforcers for appropriate staff behavior which related to dental hygiene. They combined three permanent product measures of resident treatment taken several times daily, with a six play check-list of staff activity, used sixteen times per workshift according to a momentary time sampling procedure (Powell, Martindale & Kulp, 1975). This assessment strategy indicated which of the targeted behaviors came under control of the reinforcers used, and which did not. A second study replicated the first and extended the assessment of patient dental care to include rated "quality," using grades assigned to technicians by the dental assistant. When eligibility for choosing workshift preferences was made contingent upon good dental care grades, the frequency of "F" grades dropped to 25% of the baseline rate of 19% (i.e., 5%), while the frequency of "A" and "B" grades rose from 35% to 65%. Hence, both activity of the staff, and the quality of that activity could be brought under experimental control.

Multiple measure assessments are better suited for conducting the comprehensive measurements required by process research. It is apparent that there are limitations inherent in these designs as well; both in terms of the diminished reliability of complex observations (Taplan & Reid, 1973) as well as the insensitivity of predefined categories to covariates such as degraded quality of service and so forth. To assess the

quality of dental care, Iwata et al. (1976) required the expensive services of an independent assessor; such resources are rarely at the disposal of a program coordinator (Reppucci & Saunders, 1974).

The relationship between the quality of staff-patient interactions has been long acknowledged. Galioni, Adams and Tallman, (1953) assessed the effects of doubling the number of patient-staff contacts on a psychiatric hospital ward. At nearly twice the cost to a regular ward, this intervention yielded only non-significant trends towards increased release rates. Hyde (1953) demonstrated that the frequency of friendly patient-patient contacts is a direct function of the frequency of friendly staff-patient contacts. This suggests that interventions (e.g., Galioni et al., 1953) that merely increase the number of staff-patient contacts are potentially trivial unless efforts are taken to ensure that the quality of these contacts is maintained at a high level. The most satisfactory way to assess the quality of change agent-target contacts is through the measurement of the interactions between these two groups.

Interaction Assessments

The most sensitive measurement of staff performance is possible only when the behavior is assessed within the ever changing environmental flux within which all activity occurs

(Patterson, 1974). Interaction measures account for the changing behavior of either party as a part of the determining environment of the other. Such measures can be used to demonstrate functional control over specified behavior patterns, and provide greater information and more substantial empirical support for the assertion of functional control than the mere correlation of two independently measured targets (cf. Hall et al., 1968).

Most reported interaction measures are rather simplified schemes in which the behavior of both the change agents and the target members are dichotomized along some relevant variable(s), which yields four possible behavior cells. Commonly, only three of these four cells are recorded, and the remaining behavior is recorded by exclusion (Cooper, Thomson, & Baer, 1970; Herbert & Baer, 1970). These authors were able to assess significant changes in both the frequency of behaviors and the interaction sequences following performance training (Herbert & Baer, 1970) or intervening by providing performance feedback (Cooper et al., 1970). In the Herbert study, two mothers self monitored in order to obtain their performance feedback, and although objective ratings indicate that their monitoring was grossly inaccurate, that had little bearing on their performance. This self monitoring exemplifies another strategy advantageous for use with clinical populations due to its reactive effects

but unsuitable for many research purposes that attempt to identify external controlling influences that contribute to observed changes.

It is possible to improve the efficiency of the assessment device still further than in the preceeding two studies. For example, Greenwood, Hops, Pelaquadri, and Guild (1974) defined and scored only three categories of teacher attention, viz., correct teacher attention (following appropriate group response), incorrect positive teacher attention (following inappropriate group response), and incorrect teacher attention (independent of group response). By utilizing a higher degree of abstraction in the design of the assessment device, these authors were able to assess three categories of teacher behavior and two categories of student behavior, all within a simple three category system. A minimum number of categories provides conditions favorable to the attainment of high reliability of observation (Kazdin, 1977; Taplan & Reid, 1973).

Although even more sophisticated assessments are possible, there is but a single report of a staff assessment device in the tradition of the sensitive child interaction scoring systems of Patterson or Wahler (Patterson & Reid, 1974; Reid, 1977; Wahler, House, & Stambaugh, 1976). This study (Parsonson, Baer, & Baer, 1974) assessed the application of "correct social attention" by two teacher's aides. Ten classes of child behavior (all but two of which) could be scored as situationally appro-

priate or inappropriate. Teacher behavior was scored only for positive or negative attention, but owing to the definitions of child appropriate and inappropriate behavior, it is a straightforward calculation to determine instances of teacher appropriate versus inappropriate responses to each instance of child behavior. This study demonstrates an effective technique for training generalized correct social attention, and deserves recognition as a landmark in the field of change agent assessment and intervention.

Conclusions

It is evident that none of the procedures reviewed are capable of identifying both gross changes (e.g., Iwata et al., 1976; Montegar, Reid, Madsen, & Ewell, 1977) and the subtle assessment of agent-target interactions (e.g., Parsonson et al., 1974). Although not presently reported in the literature, a more comprehensive assessment of change agent performance becomes possible when an interaction measure is embedded into those categories of a multiple measure scoring device which identify instances of agent-target contact. An outline of such a procedure is attached in Appendix I for comments and criticisms before the device is piloted. A further suggestion accompanies the welcome reintroduction of expectancies and minds into the mainstream of psychology. This is the role of attitudes of change agents, and their relationships to therapeutic

outcomes. Gripp and Magaro (1974) identify attitude as potentially the most important determiner of change agent performance, and hence of therapeutic success. There is much need for research on noncoercive techniques to assess and implement attitude changes.

What common conclusions can be gleaned from the studies reviewed? For the purposes of intervening via the provision of feedback, there appears to be little value in reporting negative instances of staff behavior in feedback sessions (Coleman & Buren, 1969), unless it is used in conjunction with feedback about positive agent behavior (Parsonson et al., 1974; Greenwood et al., 1974). Even then, Cooper (Cooper et al., 1970) asserts that if negative feedback is to be used at all, it should target missed opportunities for correct behavior, rather than incidents of incorrect staff behavior. Considerations of patient care follow below.

There are several reasons to favor group over individual interventions in many cases. Group programs favor a more equitable distribution of change agent attention to all group members (Manelker et al., 1970). Group programs tend to cost less to implement (Jones & Eimers, 1975), and they tend to place more reliance on naturally occurring social contingencies to control behavior (Baer & Wolf, 1970). There is further evidence that behavior is under greater control by status

level reinforcement, in which inadequate performance on any of several behavioral categories is just cause to produce a demotion of status (Coleman & Buren, 1969; Phillips, 1968).

Much has been accomplished towards identifying variables which exert influence over staff performance (cf. Atthowe & Krassner, 1968). The need to program generalization into training programs is documented by Parsonson et al. (1974) and others Horton, 1975; Miller & Sloane, 1976). The promise of tangible reward has been shown to better facilitate arbitrary button-presses by change agents than did information that related those responses to reports of patient improvement (Loeber, 1971), but it is not at all clear what relationship exists between experimental button presses and typical patient care activities. Potential reinforcers available to program coordinators include coffee fund exclusions, break time or work shift preferences, day off preferences, social approval, recognition, or eligibility to win prizes or privileges in a lottery.

The question of change agent performance attempts to develop ways to maximize the effectiveness of any given program. Gripp and Margaro (1974) propose that there are often unrecognized and uncontrollable feedback loops that may be responsible for many changes observed following program introduction. Assessment of changes in staff performance helps to identify the concurrent changes in behavior among members of target

populations and the change agents themselves. This information can suggest specific factors which are directly and indirectly responsible for producing therapeutic changes.

References

Andrasik, F., & McNamara, J. Optimizing staff performance in an institutional behavior change setting: A pilot study. Behavior Modification, 1977, 1, 235-248.

Atthowe, J., & Krassner, L. Preliminary report on the application of contingent reinforcement procedure (token economy) on a "chronic" psychiatric ward. Journal of Abnormal Psychology, 1968, 73, 37-43.

Bandura, A. Social Learning Theory. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Barnard, J., Christophersen, E., & Wolf, M. Supervising paraprofessional tutors in a remedial reading program. Journal of Applied Behavior Analysis, 1974, 7, 481. (abstract)

Bensberg, G., & Barnett, C. Attendant training in southern residential facilities for the mentally retarded. Atlanta: Southern Regional Educational Board, 1966.

Berkowitz, B., & Graziano, A. Training parents as behavior therapists. Behavior Research and Therapy, 1972, 10, 297-317.

Broden, M., Bruce, C., Mitchell, M., Carter, V., & Hall, R. Effects of teacher attention on attending behavior of two boys at adjacent desks. Journal of Applied Behavior Analysis, 1970, 3, 199-203.

Bushell, D., Wrobel, P., & Michaelis, M. Applying group contingencies to the classroom study behavior of preschool children. Journal of Applied Behavior Analysis, 1968, 1, 55-61.

Coleman, A., & Buren, J. An information system for measuring patient behavior and its use by staff. Journal of Applied Behavior Analysis, 1969, 2, 207-214.

Cooper, M., Thomson, C., & Baer, D. The experimental modification of teacher attending behavior. Journal of Applied Behavior Analysis, 1970, 3, 153-157.

Craighead, E., Kazdin, A., & Mahoney, M. Behavior Modification. Boston: Houghton Mifflin, 1976.

Daily, W., Allem, G., Chinsky, J., & Veit, S. Attendant behavior and attitudes towards institutionalized children. American Journal of Mental Deficiency, 1974, 78, 586-591.

Fielding, L., Errickson, E., & Bettin, B. Modification of staff behavior: A brief note. Behavior Therapy, 1971, 2, 550-553.

Galioni, E., Adams, F., & Tallman, F. Intensive treatment of backward patients: A controlled pilot study. American Journal of Psychiatry, 1953, 109, 576-583.

Gardner, J. Teaching behavior modification to nonprofessionals. Journal of Applied Behavior Analysis, 1972, 5, 517-522.

Gardner, J. Training the trainers: A review of research on teaching behavior modification. In R. Rubin, J. Brady, & J. Henderson (Eds.) Advances in Behavior Therapy. New York: Academic Press, 1973. Pp. 145-176.

Greenwood, C., Hops, H., Delquadri, J., & Guild, J. Group contingencies for group consequences in classroom management: A further analysis. Journal of Applied Behavior Analysis, 1974, 7, 413-425.

Gripp, R., & Magaro, P. The token economy program in the psychiatric hospital: A review and analysis. Behavior Research and Therapy, 1974. 12, 205-228.

Harmatz, M. Observational study of ward staff behavior. Exceptional Children, 1973, 39, 554-558.

Herbert, E., & Baer, D. Training parents as behavior modifiers: Self recording of contingent attention. Journal of Applied Behavior Analysis, 1972, 5, 139-150.

Horton, G. Generalization of teacher behavior as a function of subject matter specific discrimination training. Journal of Applied Behavior Analysis, 1975, 8, 311-319.

Hyde, R. Factors in group motivation in a mental hospital. Journal of Nervous and Mental Disorders, 1953, 117, 212-225.

Iwata, B., Bailey, J., Brown, K., Foshee, T., & Alpern, M. A performance based lottery to improve residential care and training by institutional staff. Journal of Applied Behavior Analysis, 1976, 9, 417-431.

Jones, F. & Eimers, R. Role playing to train elementary teachers to use a classroom management "skill package." Journal of Applied Behavior Analysis, 1975, 8, 421-433.

Kazdin, A. Artifact, bias and complexity of assessment: The ABC's of reliability. Journal of Applied Behavior Analysis, 1977, 10, 141-150.

Kazdin, A. Methodological and assessment considerations in evaluating reinforcement programs in applied settings. Journal of Applied Behavior Analysis, 1973, 6, 517-531.

Manelker, A., Brigham, T., & Bushell, D. The effects of token procedures on a teacher's social contacts with her students. Journal of Applied Behavior Analysis, 1970, 3, 169.

Milby, J. A review of token economy treatment programs for psychiatric inpatients. Hospital and Community Psychiatry, 1975, 26, 651-658.

Miller, S., & Sloane, H. Generalization of parent training. Journal of Applied Behavior Analysis, 1976, 9, 355-370.

Montegar, C., Reid, D., Madsen, C., & Ewell, M. Increasing institutional staff to resident interactions through inservice training and supervisor approval. Behavior Therapy, 1977, 8, 533-540.

Nay, W. A systematic comparison of instructional techniques for parents. Behavior Therapy, 1975, 6, 14-21.

O'Dell, S. Training parents in behavior modification: A review. Psychological Bulletin, 1974, 7, 418-433.

Parsonson, B., Baer, A. & Baer, D. The application of generalized correct social contingencies. An evaluation of a training program. Journal of Applied Behavior Analysis, 1974, 7, 427-437.

Patterson, G. A basis for identifying stimuli which control behaviors in natural settings. Child Development, 1974, 45, 900-901.

Powell, J., Martindale, A., and Kulp, S. An evaluation of time-sample measures of behavior. Journal of Applied Behavior Analysis, 1975, 8, 463-469.

Reid, J. A social learning approach to the treatment and study of families: II. An observational system for the collection of observation data in the home setting. Eugene, Oregon: Castilia Publishing, 1977. (in press)

Loeber, R. Engineering the behavioral engineer. Journal of Applied Behavior Analysis, 1971, 4, 321-326.

Phillips, E. Achievement Place: Token reinforcement in a home-style rehabilitation setting for "pre-delinquent" boys. Journal of Applied Behavior Analysis, 1968, 1, 213-223.

Reppucci, N., & Saunders, J. Social psychology of behavior modification: Problems of implementation in natural settings. American Psychologist, 1974, 29, 649-660.

Skinner, B. Beyond Freedom and Dignity. New York: Alfred A. Knopf, 1971.

Taplan, J., & Reid, J. Effects of instructional set and experimenter influence on observer reliability. Child Development, 1973, 44, 547-554.

Tharp, R., & Wetzel, R. Behavior Modification in the Natural Environment. New York: Academic Press, 1969.

Wahler, R. Some structural aspects of deviant child behavior. Journal of Applied Behavior Analysis, 1974, 7, 151-165.

Wahler, R., House, A., & Stambaugh, E. Ecological Assessment of Child Problem Behavior. New York: Permagon Press, 1976.