Measurement issues relating to behavior modeling research are discussed. Researchers have successfully demonstrated the ability of behavior modeling to produce significant behavioral change in the training milieu, but have been less successful in showing the transfer of such change to everyday situations, a problem partially caused by measurement deficiencies. Comparison of pre-training and post-training role playing is a common basis for assessment. Role-playing success may have potential for a high level of ecological validity, but this is not ascertainable unless assessment is based on multiple samples of behavior. Other problems commonly associated with behavior modeling research include: (1) inadequate behavior sampling; (2) insufficient consideration of lagged effects; and (3) inadequate attention to the problem of proactive inhibition. A four-page list of references is included. (SLD)
Some Measurement Issues Relating to Behavior Modeling

William M. Fox
Department of Management
College of Business Administration
University of Florida
Gainesville, Florida 32611

Presented at the annual meeting of the American Psychological Association, New York, August 1987.
Some Measurement Issues Relating to Behavior Modeling

Research

Abstract

Roleplays may provide an acceptable means for assessing behavioral change, but only when they are properly used. The following deficiencies are commonly associated with Behavior Modeling research: inadequate behavior sampling, insufficient consideration of lagged effects, and inadequate attention to the problem of proactive inhibition. Each of these is discussed in relation to relevant research.
Some Measurement Issues Relating
to Behavior Modeling Research

The basic Behavior Modeling procedure involves the following elements: 1) instructions for mastering a skill 2) demonstration of the proper performance of the skill by a competent model 3) overt trainee practice of the demonstrated behavior 4) feedback to trainee about performance, and 5) follow-up to encourage sustained mastery by the trainee and transfer of the skill to non-training milieus.

Researchers have been highly successful in demonstrating the ability of Behavior Modeling to produce significant behavioral change in the training milieu. They have been less successful in demonstrating transfer of such change to everyday situations (for recent evidence to this effect, see meta-analysis by Falcone, Edwards, and Day, 1986).

Undoubtedly, part of the problem deals with the effective management of the transfer process: provision of effective behavior management in an environment that is supportive of the new behaviors. However, it is clear that measurement deficiencies play an important role, also. The purpose of this paper is to explore the latter aspect of the problem.

Roleplays as a Measure

Comparison of pre-training and post-training role-playing performances provided the most common "objective" basis for assessing training effects in a majority of the 39 single-skill-area studies reviewed by the author. Does this mean that the trainees had simply learned to be better role-players?: a legitimate concern explored by McNamara and Blumer (1982).
One aspect of this question deals with whether or not practice, per se, improves role-playing performance. Five studies from the present survey provide data which relate to this issue (Hersen, Eisler, Miller, Johnson, & Pinkston, 1973; Melnick, 1973; Eisler, Hersen, and Miller, 1973; McFall & Twentyman, 1973, Experiments 1 and 3). In all five, post-test role-playing scores did not significantly exceed pre-test scores for the controls who did not receive training. Additionally, in two of the studies (Hersen et al., 1973; Eisler et al., 1973), four separate roleplays in practice-only control groups failed to produce significantly higher post-test role-playing scores.

Clearly, the positive improvements in role-playing performance demonstrated by the trainees in these studies were not the result of practice, alone. Roleplays, then, may provide an appropriate medium for assessing behavioral change, but have they been used properly?

Need For Sound Behavior Sampling

The inadequacies of self-report data have been well documented via comparisons with more objective behavioral assessment data. For examples, see the studies by Wolfe and Fodor (1977), Kazdin (1974), and McFall and Lillesand (1971).

The equivocal results obtained with single follow-up telephone calls (see McFall & Galbraith, 1977; Thorpe, 1973; Kazdin, 1974) or single observation periods (see Bouffard, 1073; Schinke & Rose, 1976; Sorcher & Spence, 1982) would be expected by Epstein (1979). He notes that correlations between single observations of expectedly similar behavior rarely exceed .30, even when taking the
interaction of the person with the situation into account, no matter how careful the measurement.

However, his studies found that the stability of self-recorded data, observations of overt behavior, and direct measurements of objective behavior increased dramatically, from very low levels for one day of observations to quite respectable levels for 5 days, tending to peak at impressive levels at about 14 days. He found that temporal reliability consistently exceeded .70 when a sufficient number of behavior samples was taken.

Epstein's findings about the inadequacies of single observations are supported by the reviews of Bowers (1973) and Endler and Magnusson (1976). They report reliability coefficients between two samples of behavior generally in the .20-.50 range.

Their reviews also establish the importance of taking into account the interaction of the person with the situation in predicting behavior. In most instances, they found that the interaction accounted for more variance than either person or situation, alone.

In turn, this would seem to explain the findings of McNamara and Blumer (1982). Their comprehensive review of studies which examine the relationship between a sample of behavior in a naturalistic setting and a single roleplay revealed moderate correlations, generally in the .30-.50 range.

It seems clear, then, that superior role-playing in a realistic training milieu has the potential for a high level of ecological validity, but this will not be ascertainable unless behavioral assessment is based upon multiple samples of behavior.
Several studies have, in effect, utilized post-training behavioral assessments which reflect a multiple sampling of behavior. Cooker and Cherchia (1976) measured the effectiveness with which trainees conducted meetings during 3 separate, one-hour sessions. Dodd and Pesci (1977) and Smith (1976, First Study) measured the effectiveness with which managers conducted extended or multiple meetings: first, to feed back and to analyze opinion survey data and, second, to prepare meaningful action plans for many of the problems which were uncovered. Porras and Anderson (1981) recorded subordinate perceptions of characteristic rates of different types of behavior 6 months after training, and Latham and Saari (1979) had managers of the supervisor trainees evaluate them on Behavior Observation Scales one year after training. Latham and Saari also used multiple role playing tests - one for each of the 7 training modules - to measure training effects.

Other studies measured factors which were presumed to be the result of behavior change. For example, Kelly, Laughlin, Claiborne, and Patterson (1979) used success in actual employment interviews to assess the effectiveness of interview training. Sarason and Ganzer (1973) compared rates of recidivism for juvenile delinquents 3 years after training in social skills. Goldstein and Sorcher (1974) used separation rates and various measures of worker productivity to assess the effects of supervisory skills training.

The Factor of Timing

The timing of post-training behavioral measurement may be of far more importance than many researchers realize. The structured group interviews of Sorcher and Spence (1982) with 3-4 subordinates of each supervisor, designed to identify specific changes in interactions between supervisors and employees, turned up
insignificant positive changes in supervisory behavior in a first post-test 6 weeks after training, but statistically significant changes in a second post-test 14 weeks later.

Burnaska (1976) found that trained managers performed better 4 months after training, than immediately after training, on blindly evaluated roleplays; yet, subordinate perceptions of their managers' behavior indicated only slight change. He speculates that 4 months may not have been enough time for managers to use their new skills with sufficient frequency to produce more perceptual change. Support for this interpretation was provided by large within-cell variances for each of the analyses of the perceptual data.

Data from a number of studies indicate, also, that positive changes in subordinate performance may lag positive changes in the behavior of their supervisors by several months to several years, with the more prevalent time frame being 7-18 months (Franklin, 1975; Likert, 1967; Likert & Likert, 1976; Marrow, Bowers & Seashore, 1967; Taylor & Bowers, 1972).

Need to Consider Training Time

A treatment aspect which is inadequately reported in many studies on Behavior Modeling is the amount of training time invested per trainee. For the 27 single-module studies which do permit an estimate, the time per trainee ranges from 2 to 450 minutes, with an average requirement of 27 minutes (excluding the one, extreme case of 450 minutes). For the 8 multiple-module studies which permit an estimate, the time per trainee per module ranges from 8 to 60 minutes, with an average of 24 minutes.
These widely varying time allocations point up an important consideration: the amount of repetition or overlearning that is required for effective retention/implementation, in contrast with that required simply for module mastery. It seems likely that inadequate attention to this matter accounts for a large proportion of the poor transfer of training results reported for many of the single-module social skills studies, despite the significant immediate post-treatment positive training effects achieved in virtually all of them.

To lay the groundwork for effective retention and transfer of learning, training of this type must deal effectively with the problem of proactive inhibition: the lasting retention of a new approach following the well-practiced establishment of an old approach. To the extent that there is such interference from a competing system, trainees will require practice beyond mere short-term mastery and retention to help assure longer-term displacement of the older system. It is in this regard that overlearning plays an important role in Behavior Modeling as is demonstrated by Saltz (1971, pp. 197, 206) and by Goldstein & Sorcher (1974, pp. 61-62, 75).

The problem of interference is greatest when two competing systems of behavior are related to the same discriminative stimulus or antecedent cue (Saltz, 1971, pp. 197-206). Clearly, this is the case with most Behavior Modeling training. Our goal, typically, is to displace less productive ways of responding to given situations with more productive ways. Researchers can greatly facilitate the
study of overlearning requirements by consistently reporting the amount of training time devoted to each trainee when the purpose of the training is to replace a well-established behavior with a new behavior.
References


*A review of this work is readily available: