

*THE CHANGING CRITERION DESIGN*DONALD P. HARTMANN<sup>1</sup> AND R. VANCE HALL

UNIVERSITY OF UTAH AND UNIVERSITY OF KANSAS

This article describes and illustrates with two case studies a relatively novel form of the multiple-baseline design called the changing criterion design. It also presents the design's formal requirements, and suggests target behaviors and circumstances for which the design might be useful.

DESCRIPTORS: multiple baseline, changing criterion design, experimental design, methodology, shaping, experimental control

The development of experimental designs to demonstrate control in individual case studies has been a crucial factor in bringing scientific status to the study of individuals. Applied behavioral research has primarily used reversal and multiple-baseline designs (Baer, Wolf, and Risley, 1968; Barlow and Hersen, 1973; Hall, Cristler, Cranston, and Tucker, 1970; Leitenberg, 1973; Wolf and Risley, 1971), although other designs suitable for individual subject research have been described by Edgington (1969, pp. 135-140), Gelfand and Hartmann (1968, p. 211), and Gottmann (1973). The advantages and limitations of these designs have been discussed by Gelfand and Hartmann (1975), Jones (1974), and McNamara and MacDonough

(1972), among others. This paper describes an additional design that might be particularly useful in individual case studies when other designs are inconvenient or unsuitable, and changes in the target behavior are made in stepwise increments.

The changing criterion design, initially named by Hall (1971) and illustrated by Weis and Hall (1971), is a variation of a multiple-baseline design and is similar to a design described, but unnamed, by Sidman (1960, pp. 254-256). The design requires initial baseline observations on a single target behavior. This baseline phase is followed by implementation of a treatment program in each of a series of treatment phases. Each treatment phase is associated with a stepwise change in criterion rate for the target behavior. Thus, each phase of the design provides a baseline for the following phase. When the rate of the target behavior changes with each stepwise change in the criterion, therapeutic change is replicated and experimental control is demonstrated.

## CASE I

The changing criterion design is illustrated with data from two case studies from our laboratories. The first study, taken from Hall and Fox (*in press*), used the changing criterion design to demonstrate experimental control over the num-

<sup>1</sup>We were independently preparing papers on the criterion change design in mid-1974. Through the *JABA* review process, we became aware of one another's work on the topic and agreed to co-author the present paper. Elements of the research and writing of this article were supported in part by National Institute of Mental Health HDMH-06914 to Donna M. Gelfand and Donald P. Hartmann, University of Utah, Salt Lake City, Utah, and by National Institute of Child Health and Human Development HD-03144 to the Bureau of Child Research, Department of Human Development and the Department of Special Education, University of Kansas. Reprints may be obtained from Donald P. Hartmann, Department of Psychology, University of Utah, Salt Lake City, Utah 84112, or from R. Vance Hall, Juniper Gardens Children's Project, 2021 North Third Street, Kansas City, Kansas 66101.

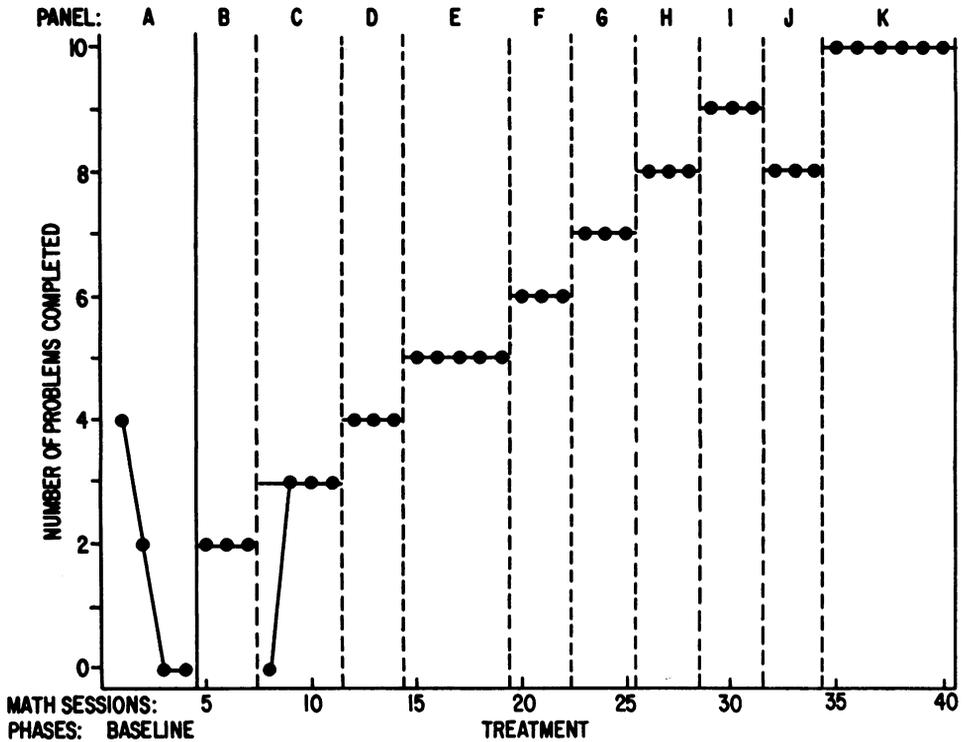


Fig. 1. A record of the number of math problems correctly solved by a behaviorally disordered boy during baseline, and when recess and the opportunity to play basketball were made contingent on changing levels of performance. Solid horizontal line segments indicate criterion for each treatment phase.

number of math problems correctly solved by a student in a classroom for behaviorally disordered children (see Figure 1). During the baseline phase, the teacher gave the student a worksheet with nine division problems. In baseline (Panel A), the number of problems completed decreased from four during the first day to zero. In the first treatment phase (Panel B), the criterion number of problems to be worked was set at two, the next highest whole problem over the baseline mean. The consequences for correctly solving two problems on the worksheet during a 45-min work session included access to recess after the session and the opportunity to play basketball. Failure to complete two problems within the allotted time prolonged math time until the problems were correctly computed. In subsequent treatment phases, identical consequences were in effect and the criterion was advanced by one problem after three consecutive days of performing at the specified level.

During the fourth treatment phase (Panel E), the problem criterion level was maintained for five rather than three consecutive days, and during the ninth treatment phase (Panel J), the criterion was dropped one problem, rather than raised one problem. Following this final worksheet phase, the subject was required to solve 10 problems in the math text correctly (Panel K) under the same contingency arrangement. During this, as well as prior treatment phases, math performance perfectly matched the criterion rate with but a single exception (Panel C).

Although these data demonstrated that the treatment package exerted control over math problem solving, they do not indicate which of the treatment components in fact exercised control. Component analysis requires additional design features, including conditions in which one or more treatment element is omitted [see, for example, Hall and Fox, Experiment III (*in press*)].

## CASE II

The second study used the changing criterion design to demonstrate experimental control over smoking rate in a deceleration program (see Figure 2). The subject's baseline rate of smoking is given in Panel A; following this baseline phase, treatment was instituted. Treatment consisted of imposing a daily criterion rate of smoking and consequences for over- and under-indulgence. In the first treatment phase (Panel B), the criterion rate was set at 95% of baseline or 46 cigarettes per day. Cigarettes smoked in excess of this rate resulted in a \$1.00 escalating fine (\$1.00 for the forty-seventh cigarette, \$2.00 for the forty-eighth cigarette, etc.). Smoking fewer than the criterion number of cigarettes resulted in a \$0.10 escalating bonus. For the second (Panel C) and subsequent treatment phases, only four of 21 actually conducted being presented here (Panels D, E, F, and G), the criterion was set at 94% of the criterion rate for the previous treatment phase; identical consequences were imposed during all treatment phases.

The underlying multiple-baseline features from which the changing criterion design derives its credibility are more clearly illustrated by regraphing the smoking study data (see Figure 3). These data consist of the percentage of days within each week that four arbitrarily chosen criteria were met.<sup>2</sup> They indicate that the percentage of days on which the yet-to-be-applied criteria are met is low (0% to 29%), while the percentage for the current and previously applied criteria is high (invariably 100%). Furthermore, stepwise changes in percentage score (from low to high) occur only as each criterion is applied.<sup>3</sup> The pattern and tim-

<sup>2</sup>Because the data in Figure 3 are based on weekly performance, rather than daily performance as in Figure 2, Figure 3 includes data from 53 weeks, rather than from the 13 weeks presented in Figure 2.

<sup>3</sup>If more closely adjacent changes in the criterion had been selected for inclusion in Figure 3, the results would have been somewhat less clear than those shown, particularly for changes in the criterion instituted early in treatment, when within-phase behavioral variability was substantial. On the other hand, selection of adjacent changes in the criterion later

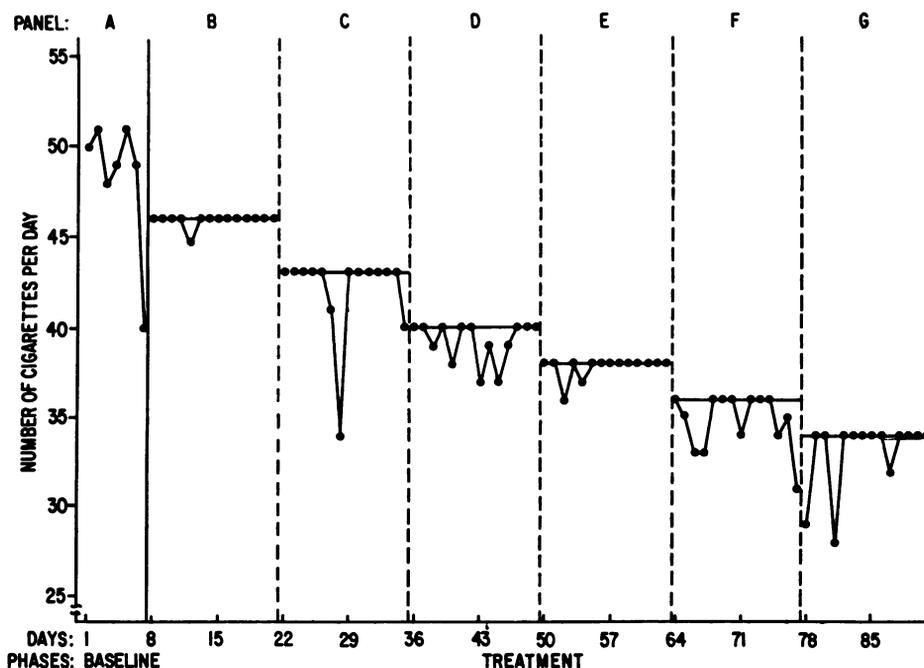


Fig. 2. Data from a smoking reduction program used to illustrate the stepwise criterion change design. Solid horizontal line segments indicate criterion for each treatment phase.

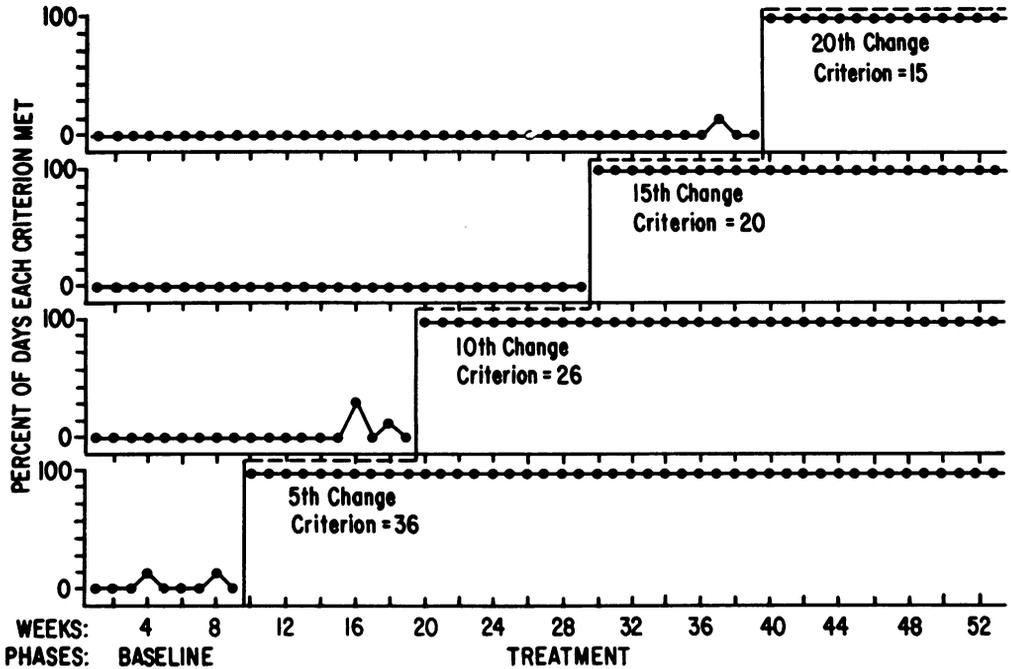


Fig. 3. Regraphing of smoking reduction program data to illustrate multiple-baseline characteristics of the changing criterion design. Note that this figure contains data from more treatment phases than does Figure 2. The data for Weeks 12 and 13 on this figure correspond to the data in Panel G of Figure 2.

ing of changes in rate resembles the pattern and timing of changes in rate expected with the successful application of a multiple-baseline design. Thus, with the changing criterion design, changes in the criterion function like the sequential changes in the behavior, situation, or person to which treatment is applied with traditional variants of the multiple-baseline design.

## DISCUSSION

Successful implementation of the changing criterion design requires particular attention to three design factors: length of baseline and treatment phases, magnitude of changes in the criterion, and number of treatment phases or changes in criterion. All phases should be long enough to ensure that successive changes in a therapeutic direction are not naturally occurring due to either historical, maturational, or mea-

surement factors (see Campbell and Stanley, 1963). In addition, treatment phases should differ in length, or if of a constant length, should be preceded by a baseline phase longer than each of the separate treatment phases. This is to ensure that stepwise changes in the rate of the target behavior are not occurring naturally in synchrony with criterion changes. The baseline data should also be stable (*e.g.*, zero slope) or should be changing in a counter-therapeutic direction. While the baseline data shown in Panel A of Figure 1 meet both of these formal design criteria, the baseline data shown in Panel A of Figure 2 unfortunately do not meet either criterion.

The length of each treatment phase and the magnitude and number of criterion changes are interdependent and should vary as a function of the total length of treatment, the variability of the target behavior, and the difference between the baseline and the anticipated terminal rate of the target behavior. Each treatment phase must be long enough to allow the rate of the

in treatment when within-phase variability was virtually zero would have produced even clearer results than those shown in Figure 3.

target behavior to restabilize at a new and changed rate; it is stability after change has been achieved, and before introduction of the next change in the criterion, that is crucial to producing a convincing demonstration of control. Thus, for example, when behavior change is slow, treatment phases must be longer to ensure stability before the criterion is again changed for the next treatment phase. In deciding the length of each treatment phase, clinician-researchers must also remain sensitive to their clients' needs and not hold clients at a criterion longer than would be wise therapeutically.

The magnitude of each stepwise change in the criterion must be large enough to ensure that the changes in the rate of the target behavior are detectable. Thus, the size of changes in the criterion employed will be largely a function of the variability of the rate of the target behavior within each treatment segment. With highly variable responses, larger changes in the criterion (and longer treatment phases) will be required to demonstrate changes in rate. Therapeutic considerations such as the difficulty of changing the rate of the problematic behavior will also determine the magnitude of changes in the criterion employed. With behavior that is difficult to change, the changes in the criterion should be small enough to be achievable, yet large enough to be clearly detectable.

Demonstration of experimental control with the changing criterion design is best shown by a visual display of the close correspondence between behavior rate and changes in the criterion. However, with data characterized by excessive variability or complex trends, a simple visual display may be unconvincing, and statistical procedures may be required. In such cases, correlational analyses (Hall and Fox, *in press*), the analysis of variance (Hartmann, 1974), or time-series analysis (Jones, Vaught, and Weinrott, *in press*) may be applied to the data from a changing criterion design. These techniques should be used with caution, however, as they either are controversial, necessitate conformity to a sub-

stantial number of statistical assumptions, or require extensive data (see, for example, Baer (*in press*); Kazdin (1976); Michael, 1974; Thoresen and Elashoff, 1974).

The number of criterion changes included is dependent on the number of replications required for a convincing demonstration of experimental control. Current practice suggests that two replications in a multiple-baseline design may be sufficient when the correspondence between changes in behavior rate and changes in condition are clear. More than two replications may be desirable when the correspondence is less clear; for example, when the rate of the target behavior regularly exceeds the criterion. When the rate of the target behavior fails to track the criteria closely, Hall and Foxx (*in press*) suggest that demonstration of control may be further strengthened by reverting temporarily either to a former criterion (see Panel J of Figure 1), or to baseline conditions (see Deitz and Repp, 1973, Experiment III). Still another technique to fortify a weak or questionable demonstration of control in a design with equal length treatment phases is to leave the criterion at an established level for a longer period than required by ordinary stability considerations (see Panel E of Figure 1).

The logic of the changing criterion design does not require counter-therapeutic behavior change, unlike the reversal design. Nor does the design require multiple independent behaviors, subjects, or situations, unlike standard multiple-baseline designs. Instead, the changing criterion design should be applicable to a wide range of treatment problems that can be modified in a stepwise manner and where reasonably prompt changes to a new, stable level are expected in response to changes in the criterion. Thus, as Hall and Fox (*in press*) suggest, the changing criterion design would be particularly suited to demonstrating the effectiveness of shaping procedures. In general, the changing criterion design should be useful for acceleration problems in which stepwise increases in accuracy, frequency, duration, latency, or magnitude are

therapeutic goals; e.g., increases in writing or reading rate, improvements in peer interactions, or better adherence to an exercise routine. Deceleration problems for which stepwise decreases in similar response characteristics are therapeutic goals include reductions in smoking, drinking, or overeating, and latency of compliance with instructions.

In summary, the changing criterion design is capable of providing convincing demonstrations of experimental control, seems applicable to a wide range of problematic behaviors, and should be a useful addition to applied individual-subject methodology.

### REFERENCES

- 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.
- Barlow, D. H. and Hersen, M. Single-case experimental designs. *Archives of General Psychiatry*, 1973, 29, 319-325.
- Campbell, D. T. and Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1963.
- Dietz, S. M. and Repp, A. C. Decreasing classroom misbehavior through the use of DRL schedules of reinforcement. *Journal of Applied Behavior Analysis*, 1973, 6, 457-463.
- Edgington, E. S. *Statistical inference: The distribution-free approach*. New York: McGraw-Hill, 1969.
- Gelfand, D. M. and Hartmann, D. P. Behavior therapy with children: A review and evaluation of research methodology. *Psychological Bulletin*, 1968, 69, 204-215.
- Gelfand, D. M. and Hartmann, D. P. *Child behavior analysis and therapy*. New York: Pergamon Press, 1975.
- Gottman, J. M. N-of-one and N-of-two research in psychotherapy. *Psychological Bulletin*, 1973, 80, 93-105.
- Hall, R. V. *Managing behavior: Behavior modification, the measurement of behavior*. Lawrence, Kansas: H & H Enterprises, 1971.
- Hall, R. V., Cristler, C., Cranston, S. S., and Tucker, B. Teachers and parents as researchers using multiple-baseline designs. *Journal of Applied Behavior Analysis*, 1970, 3, 247-255.
- Hall, R. V. and Fox, R. G. Changing criterion designs: An alternative applied behavior analysis procedure. In B. C. Etzel, J. M. LeBlanc, and D. M. Baer (Eds.), *New developments in behavioral research: theory method and application. In honor of Sidney W. Bijou*. Hillsdale, N.J.: Erlbaum, (*in press*).
- Hartmann, D. P. Forcing square pegs into round holes: some comments on "An analysis-of-variance model for the intrasubject replication design". *Journal of Applied Behavior Analysis*, 1974, 7, 635-638.
- Jones, R. R. Design and analysis problems in program evaluation. In P. O. Davidson, F. W. Clark, and L. A. Hamerlynck (Eds.), *Evaluation of behavioral programs*. Champaign, Ill.: Research Press, 1974.
- Jones, R. R., Vaught, R. S., and Weinrott, M. Time-series analysis in operant research. *Journal of Applied Behavior Analysis*, (*in press*).
- Kazdin, A. E. Statistical analyses for single-case experimental designs. In M. Hersen and D. H. Barlow (Eds.), *Single case experimental designs: Strategies for studying behavior change*. Oxford: Pergamon, 1976.
- Leitenberg, H. The use of single-case methodology in psychotherapy research. *Journal of Abnormal Psychology*, 1973, 82, 87-101.
- McNamara, J. R. and MacDonough, T. S. Some methodological considerations in the design and implementation of behavior therapy research. *Behavior Therapy*, 1972, 3, 361-379.
- Michael, J. Statistical inference for individual organism research: mixed blessing or curse? *Journal of Applied Behavior Analysis*, 1974, 7, 647-653.
- Sidman, M. *Tactics of scientific research*. New York: Basic Books, 1960.
- Thoresen, C. E. and Elashoff, J. D. "An analysis-of-variance model for intrasubject replication design": some additional comments. *Journal of Applied Behavior Analysis*, 1974, 7, 639-641.
- Weis, L. and Hall, R. V. Modification of cigarette smoking through avoidance of punishment. In R. V. Hall (Ed.), *Managing behavior: Behavior modification applications in school and home*. Lawrence, Kansas: H & H Enterprises, 1971.
- Wolf, M. M. and Risley, T. R. Reinforcement: Applied research. In R. Glaser (Ed.), *The nature of reinforcement*. New York: Academic Press, 1971.

Received 16 December 1975.

(Final acceptance 22 March 1976.)