INVITED BOOK REVIEW OF
SINGLE-CASE EXPERIMENTAL DESIGNS:
STRATEGIES FOR STUDYING BEHAVIOR CHANGE
BY MICHEL HERSEN AND DAVID H. BARLOW

RICHARD R. JONES

EVALUATION RESEARCH GROUP

The reviewer, Richard R. Jones, is Research Associate and President of Evaluation Research Group, Inc., 220 East 11th Avenue, Eugene, Oregon 97401, which he recently started. After receiving his PhD in 1965 from the University of Oregon, he worked as a research scientist for 11 years at the Oregon Research Institute. Jones coauthored (with John B. Reid and Gerald R. Patterson) the chapter "Naturalistic Observation in Clinical Assessment", which appeared in P. McReynolds' *Advances in Psychological Assessment*. He has written other chapters on evaluation and methodology and has contributed articles on measurement to professional journals. He is currently interested in the evaluation of behavioral applications to social problems. —EDITOR

This book, *Single-case experimental designs: strategies for studying behavior change* by Hersen and Barlow, will effectively serve its purpose, which is to provide a source book of single-case experimental design strategies, including guidelines for their use in applied research, statements of the basic goals in single-case research, and procedures required for successful execution of single-case designs. This book should be carefully studied by students of applied behavior analysis and would benefit researchers and practitioners in other areas of psychology as well. Its presentation of single-case experimental designs provides ample illustrations, shows strengths, weaknesses, advantages and disadvantages, highlights basic issues in research strategies, maintains a relatively unbiased tolerance for nonsingle-case designs, and is written in a clear and concise style.

Beyond the detailed description of single-case designs, Hersen and Barlow discuss some central issues in the design of research into clinical, behavioral disorders. For this reviewer, and seemingly for the authors as well, the paramount issue concerns the generality of findings from single-case experiments. Neither this book,

nor any other single book, will resolve the complicated design matters surrounding the generality issue. Only with the gradual accrual of research data over time will the generality issue be satisfactorily addressed. But, for now the authors provide the concerned reader with considerable food for thought about the generality-of-findings issue. In fact, this reviewer considers this issue sufficiently important to recommend that the first chapter to be read might well be the final chapter on replication, which is at the heart of the generality issue. There are several reasons. First, much of the polemic surrounding the single-case *versus* group design issue in behavioral research involves the generality-of-findings issue. Second, behavioral interventions designed to ameliorate clinical or behavioral disorders experienced a heyday in the 1960s and into the 1970s, but as the 1980s are approached, concern with the universality of these methods is becoming more and more important to practitioners, clients, and researchers alike. And third, for heuristic purposes, students and younger researchers should be informed that an issue exists. While single-case designs nicely handle many behavioral research problems, gen-

erality of findings is as much a problem for this strategy as for others. The point is that single-case researchers should be better informed about group research strategies, and traditional group researchers should be better informed about single-case experimental strategies. The generality issue is a focal point of concern to both camps and its ultimate resolution will depend not on one camp out-slugging the other, but rather on some satisfactory blending of the best aspects of both strategies, perhaps supplemented by some new research strategies yet to evolve. The generality issue will then be more satisfactorily dealt with than if strategists continue to favor one approach to the virtual neglect of the other. It is unfortunate when research approaches are discussed in an either/or fashion, *e.g.*, single-case *versus* group designs. Hersen and Barlow do well to moderate the polemic by discussing the advantages and disadvantages of single-case and group designs in a fair, albeit unbalanced manner. The imbalance is appropriately due to the focus of the book on single-case designs, not group designs. That is fine, and readers wanting more information about group designs will find references to guide their further study.

Single-case experimental designs is primarily, if not exclusively, concerned with behavioral change in individual subjects over time. The discussion of single-case experimental strategies is therefore totally appropriate, but some readers might discern a mild castigation of group research strategies because they do not function well in the study of the individual case. The relative utility of single-case and group research strategies must be considered in light of the clinical or behavioral phenomena to which the research strategy is applied. Simply stated, some research problems are best studied with single-case experimental strategies; others are best studied with group designs. Group design researchers have no more justification for criticizing single-case experimentalists concerned with individual behavioral phenomena than the reverse, since each deals with different substance

and each (hopefully) applies appropriate and suitable methodology.

If the reader accepts this view, then the parts of Hersen and Barlow's book which tend to "set up" one kind of design in opposition to the other could be bypassed quickly. The meat of the book is in two parts. First is the description of single-case experimental designs, their advantages, disadvantages, strengths and weaknesses, with ample illustrative studies. Second is the generality issue which has already been mentioned. With that perspective in mind, consider the book chapter by chapter.

Chapter 1 provides a comprehensive historical account of methodologies "over the years", with classic studies and major contributions noted. This is a good overview of clinical research strategies, although design traditionalists might feel that it critiques group designs using points that also could apply to other kinds of designs, including single-case strategies. In Chapter 2, several issues such as variability might be more understandable to the neophyte reader if basic single-subject designs are studied first, particularly since several issues in Chapter 2 are defined and then illustrated by specific designs that the neophyte reader would not have yet encountered. So, if the reader has little familiarity with single-case experimental strategies, s/he might try reading Chapter 5 before Chapter 2.

Parenthetically, an example of a questionable argument about single-subject *versus* group designs appears in Chapter 2. Hersen and Barlow suggest that one has difficulty generalizing from a group finding to an individual case. To wit, ". . . , one of the pitfalls of a truly random sample in applied research is that the more adequate the sample, in that all relevant population characteristics are represented, the less relevance will this finding have for a specific individual" (p. 55). Random sampling in psychological research is a technique designed to allow for generalization to some larger population from which the random sample has been selected. The intent of random sampling is not

to apply the findings to a specific individual. Hence, to denigrate random sampling, a group research technique, because results may not apply to a specific individual is, for this reviewer, questionable. It is not clear why one would ever go to the cost of studying a group with the intent of generalizing the findings to an individual. Common sense would say to study the individual in the first place, unless one intends to generalize the findings from the group to many individuals, *e.g.*, a population of individuals. Then, the mean of the study group is a suitable basis for generalizing to the nonstudy group (given random sampling), even though the mean individual may not represent any actual person at all.

The final section in Chapter 2 deals with the either/or polemic surrounding single-case *versus* group research strategies. Hersen and Barlow devote several pages to research problems where the group strategy seems to be the better approach and, in introducing this section, acknowledge that the critical question (and I agree) is "under what condition is one type of design more appropriate than another?" (p. 63). Researchers should not lose sight of the clear possibility that some research questions can be better dealt with by group designs than by single-case approaches, and *vice versa*.

Chapter 3 clearly describes general procedures to be used in single-case research. Several important basic rules are discussed, including the dictum that only one behavior at a time should be changed, that one cannot compare nonadjacent phases, and that equivalent numbers of data points among phases are desirable. In addition, discussions about baseline stability and suitable numbers of data points are provided. On these latter points, some research purists might be troubled by the indefiniteness of recommendations regarding matters of baseline stability or suitable numbers of data points, since the applied behaviorist must make judgement calls about each of these matters and these decisions are frequently based on the appearance of the data as the experiment unfolds. But, in the

early stages of studying a behavioral intervention, flexibility in extending baselines, varying numbers of data points, *etc.*, is seen as an advantage by applied behavioral analysts. These advantages may or may not be disadvantages from the standpoint of generality of findings. Formally, as noted in the latter part of the book, a series of replications would accommodate *ad hoc* adjustments that might be made by researchers, in that studies later in the series would become more standardized than those early in the series, where innovation might have been necessary.

Another interesting point in general research strategy occurs in Chapter 3, in the discussion of withdrawal designs. Here, as Hersen and Barlow correctly note, practical matters frequently dominate design decisions. For some, however, it may appear that these practically-based decisions may become excuses for not doing other things, such as running longer baselines or not maintaining withdrawal phases long enough really to demonstrate experimental control. In effect, some practical considerations may too easily be turned into rationales for conducting studies with less data than some would want. Too much flexibility in design matters and too much influence of practical considerations may turn some experiments into very idiosyncratic events, which would hinder replicability and ultimately the generality of findings. But, despite these few concerns, Chapter 3 is an excellent presentation of general procedures and is a chapter in this book that should be carefully studied by both new and experienced single-case researchers.

If there is any single chapter that might have been usefully expanded, it is Chapter 4. However, this reviewer would agree with the authors that measurement issues and techniques could well occupy several books. But, since repeated measurement is such a crucial component of single-case experimental strategies, sacrificing some illustrations to allow more pages for measurement issues might have been desirable. In particular, discussions of reliability and ac-

curacy could have been expanded. Fortunately, however, Hersen and Barlow provide a satisfactory literature overview so that the seriously interested reader can find the source material for self-study of these important repeated measurement issues and techniques.

In the discussion of practical considerations relating to indirect and direct systems of measurement, it is concluded that indirect techniques of measurement have limited utility in single-case research. While probably true, the tone or implication of this point is somewhat misleading. Indirect measurements include instruments such as the Rorschach, TAT, or MMPI. Whatever strengths or weaknesses these kinds of clinical or individual differences measures may possess, they were not intended to be used in frequently repeated measurement situations, as would occur in most single-case experimental designs. Yet, Hersen and Barlow imply that these measurement procedures are not useful because they do not perform well in repeated-measurement research situations. They were never intended to be used in that manner and therefore the criticism seems unjustified.

The final section in Chapter 4, dealing with kinds of measurement techniques, instruments, and procedures, provides adequate illustrations but was not intended as an exhaustive cataloguing. While the authors note this, readers should be aware that additional techniques, instruments, and procedures might be sought out for research problems that are not similar to those used as illustrations in the book.

Chapters 5, 6, and 7 carefully describe and illustrate basic, modified, and extended single-case research designs. Readers will find much detail about these designs and their applications, which is highly recommended reading for researchers faced with the choice of a design for a particular single-case research topic. For this reviewer, the most important paragraph in the book for the fledgling researcher appears on page 179, in which it is noted that the ideal conditions or procedural rules for implementation of single-case designs are rarely approxi-

mated when actually conducting research. The message for the new researcher is simply that inflexible adherence to a research design will sometimes be foolhardy because changes in procedure, design, measurement, *etc.*, may be required before a research topic or clinical problem can be satisfactorily investigated. While this point is important for users of single-case research strategies, purists might be alarmed by the flexibility needed in designs to accommodate uncooperative incoming data points. This point was made earlier in the review and bears repeating in this context. At the very least, if original plans for a study are altered "as you go along" in response to unexpected data points, perhaps the whole study should be restarted, or at least the data that caused the changes in design might be discarded. This would mean that the experimenter would "start" the study at the point in time when the data settled down.

Chapter 6 contains detailed discussion of drug study designs, with illustrations and limitations of each adequately noted. Drug research has been adequately conducted with single-case methodologies, but this is one content area where group designs, in particular factorial designs, might prove useful as well. One could question the economy of single-subject designs for component drug studies. Factorial models could be better because the generality of single-case findings becomes difficult to demonstrate as the number of interactions and combinations increases.

Chapter 7 continues the discussion of variations in single-case designs, including concurrent schedule and multiple schedule designs. These two designs are less frequently encountered in the literature, and the illustrations used in the book are less convincing than those for other designs. These illustrations may be less useful due to the infrequent use of these designs, and hence examples from which good ones could be drawn are less available.

Chapter 8, guest written by Alan E. Kazdin, covers statistical analyses for single-case experimental designs. The chapter is intended as a

nontechnical description and evaluation of different statistical tests that can be applied to intrasubject designs for the single case. The chapter accomplishes that purpose satisfactorily and fairly, given the controversy that has developed recently regarding statistics in applied behavioral research. Some readers will have to study this chapter more carefully than others in the book, but this reviewer sees that as a desirable expenditure of time and energy. The reader should be forewarned, however, that, as elsewhere in the book, there are typographical errors that can be problematic. Perhaps the major typo appears on page 279 where one reads, "a change in level across phases does not necessarily entail a change in level." This reviewer thinks it should have read, "a change in mean across phases does not necessarily entail a change in level" (or *vice versa*). The discussion of time-series analysis is a fair statement of the strengths and weaknesses of this technique that has been suggested as a supplement to visual analysis in applied behavioral research. Kazdin's presentations of randomization designs, the R_n statistic, and White's split-middle technique are complete, but some readers will require two or more readings before the concepts and details are well understood.

On the issue of whether or not statistical tests should become part of the applied behavioral researcher's armamentarium, Kazdin seems to feel that such statistical tests will be more frequently used in the future. For this reviewer, the matter boils down to whether or not statistics prove to be a useful research tool. In effect, are they functional? As of 1977, many applied behavioral researchers do not see them as functional at all. The usefulness of statistics will be the main determiner of acceptance by the field. Ideological resistance that may now exist will vanish with time as a new generation of applied behavioral researcher comes to dominate the field.

The final chapter, entitled "Beyond the Individual: Replication Procedures", was used to

introduce this review. The importance of Chapter 9, for this reviewer, is contained in the following recommendation. Any student or neophyte investigator who is planning a career in applied behavioral research should carefully study Chapter 9. Replication is clearly a canon of applied behavioral science, and is discussed frequently, but executed less frequently. Despite the exemplar series of studies described by Hersen and Barlow as illustrative of replications in applied behavior analysis, this reviewer is concerned that replications may be labelled as such, but in fact not be replications. Absolutely pure replication probably happens seldom, if ever. Pure replication would require a point-by-point duplication of a research design, varying nothing except the time the study was conducted. Such replication is considered trivial by most researchers and probably unpublishable as well. When behavioral interventions lead to large and dramatic effects, and there is no question about the experimental control demonstrated in the study, then such pure replication *is* trivial. But, when researchers change procedures (the inherent flexibility of single-case designs), plan to use the technique with different kinds of subjects in different settings, or anticipate changing any salient aspect of the design, then pure replication, of course, is impossible. Replication then becomes more a matter of repeating the work with systematic modifications. Modified procedures, subject populations, measurement systems, *etc.*, are tested to see if comparable results occur. The value of replication in single-case experimentation occurs when there is a substantial accumulation of parallel or convergent findings from a set of similar, but not identical, procedures, techniques, measurement devices, subject samples, *etc.* In the end, convergence among results from many such replications determines the generality of findings. This is the big goal to be achieved by the field of applied behavior analysis. Hersen and Barlow's book should definitely help accomplish that goal.