Designs for Evaluating Behavior Change

Conceptual Principles of Single Case Methodology

FRANK M. GRESHAM

Introduction

Documenting and evaluating the efficacy of procedures to change behavior is a hallmark of behavior therapy and applied behavior analysis. Psychologists trained in behaviorally oriented programs are taught to empirically evaluate their treatments and make treatment decisions based on empirical data. This is not to say that all behavior therapists function as researchers. Malott (1992) suggested that there are two categories of behavior analysts based on the goals and requirements of one's job: researchers and practitioners. Researchers are primarily interested in manipulating independent variables and observing their unique effects on dependent variables using tight experimental control. Practitioners are more interested in implementing a treatment program and bringing the dependent variable into acceptable ranges very quickly rather than isolating the unique effects of an independent variable. These two goals, while differing in detail and methodological rigor, are based on the same principle: data based, empirical evaluation of treatment effects or outcomes. Some may allude to the former as “applied research” and to the latter as “program evaluation.” Most behavior therapists/analysts would probably agree that collecting data on treatment outcomes is one of the most important aspects of the entire behavior change process.

This view, however, is not held by most practicing psychotherapists. A survey of clinical practitioners in the American Psychological Association by Morrow-Bradley and Elliott (1986) concerning the use of psychotherapy research by practicing psychotherapists indicated that questions addressed in research often are not clinically relevant; the variables studied are not representative of typical clinical practice; the forms in which results are reported (e.g., mean differences, $F$ ratios) do not represent clinically important changes or differences; single case research is infrequent; and practical or relevant measures of psychological change often are not used. In short, practicing psychotherapists are relatively unaffected by psychotherapy research and do not find it useful in their daily practice.

Why would these practitioners hold such a low opinion of behavior change research? One explanation may be that practicing psychotherapists, unlike behavior therapists/analysts, are not trained in a research tradition that values empirically based evaluation of treatment outcomes. In fact, many of the issues addressed in conventional psychotherapy are often not amenable to measurement and empirical evaluation. Another explanation is that the way research is presented, packaged, and marketed is often alien, incomprehensible, and irrelevant, speaking little to the practical significance of research findings. D. Barlow (1981) suggests in his criticisms of psychotherapy research that perhaps the literature

FRANK M. GRESHAM • School of Education, University of California–Riverside, Riverside, California 92521-0102.
itself, rather than the practitioner, is to blame for the underutilization of clinical research.

The purpose of the present chapter is to present readers with fundamental conceptual principles governing the planning, design, and evaluation of behavioral intervention procedures. This chapter focuses exclusively on single case methodologies used in documenting the efficacy of intervention procedures. Not all child behavior therapists or researchers adhere exclusively to a single case approach to research; however, this approach is the preferred, if not exclusive, experimental design methodology of modern-day applied behavior analysts. Given our emphasis in this book on behavior analysis, we will focus on the use of within-subject, single case experimental designs.

Overview of Research Paradigms

Two general paradigms of experimental research exist in psychology. The first approach, which is by far the most common, uses group or nomothetic data to make statements about average differences between two groups or among several groups. The second approach uses repeated measures of behavior on a single individual or on a small number of individuals to make statements about differences between control and experimental phases or conditions of an experiment. These two research paradigms differ in their goals and strategies and in the inferences made about the effects of independent variables on dependent variables.

Baer (1977) argues that the group design paradigm defends itself against spurious findings or “chance” by using samples to estimate whether or not there are population mean differences. This is accomplished by randomly sampling from a population or randomly assigning subjects to groups; these procedures, however, are not the same. What is typically done in this paradigm is to ignore this problem and assume that the difference between “A” and “B” conditions is zero.

In single case designs, defense against chance is accomplished by observing repeated patterns of behavior under repeated and alternated “A” and “B” conditions. If behavior under the repeated “A” conditions is repetitively different under the repeated “B” conditions, then it is concluded that these differences were not due to chance. These differences, however, are typically not subjected to statistical analyses to determine what degree of confidence researchers have in their findings (e.g., $p < .05$ or .01).

Researchers operating out of the group paradigm guard against errors in statistical decisions by controlling Type I errors (false rejection of a true null hypothesis). This error rate traditionally has been set at .05, thereby indicating that if a researcher concludes that the null hypothesis is false, then there is a 5% chance that this conclusion would be erroneous. Cohen (1994), a well-respected group design methodologist and statistician, argued convincingly against the use of traditional null hypothesis significance testing (NHST). He suggested that this approach has failed in advancing psychological knowledge and has actually impeded it. Cohen indicates that the chief problem in NHST is that it does not tell us what we want to know. What we want to know is, “Based on these data, what is the probability the null hypothesis is true?” In terms of a treatment outcome study involving two groups (experimental and control), what is the probability that there are not differences in the groups’ mean scores. Unfortunately, what NHST tells us is, “If the null hypothesis is true, what is the probability that these data occurred by chance?” (a Type I error). These two questions, however, are not the same.