American Journal of Community Psychology, Vol. 21, No. 5, 1993

Probing the Effects of Individual Components in Multiple Component Prevention Programs

Stephen G. West,[1] Leona S. Aiken, and Michael Todd

Arizona State University

Assessing the contributions of individual components in multi-component interventions poses complex challenges for prevention researchers. We review the strengths and weaknesses of designs and analyses that may be useful in answering three questions: (1) Is each of the individual components contributing to the outcome? (2) Is the program optimal? and (3), Through what processes are the components of the program achieving their effects? Factorial and fractional factorial designs in which a systematically selected portion of all possible treatment combinations is implemented are used to address question 1. Response surface designs in which each component is quantitatively scaled are explored in relation to question 2. Mediational analysis, a hybrid of experimental and correlational approaches, is considered in relation to question 3. Design enhancements are offered that may further strengthen some of these techniques. These techniques offer promise of enhancing both the basic science and applied science contributions of prevention research.

KEY WORDS: prevention research; multiple component intervention; mediational analysis; optimal design.

INTRODUCTION

Recent reviews have called for multiple component interventions to prevent complex social, health, and mental health problems (Hawkins, Catalano, & Miller, 1992; Shaffer, Phillips, & Enzer, 1989; Weissberg, Caplan, & Sivo, 1989). These reviews have highlighted the multiple pathways that exist between early risk factors and the later development of significant problems. Responding to these calls, recent prevention programs have included mul-

[1]All correspondence should be addressed to Stephen G. West, Department of Psychology, Arizona State University, Tempe, Arizona 85287-1104.

0091-0562/93/1000-0571$07.00/0

tiple components, each of which is targeted toward one or more of the pathways believed to underlie the development of the problem. To cite but two examples, Flay (1985) reviewed school-based social influences smoking prevention programs and concluded that the successful programs were of extended duration and were characterized by six components: media material with similar age peers, information on immediate physiological effects of smoking, correction of misperceptions about the prevalence of smoking, discussion of family and media influences on smoking and methods of dealing with them, explicit learning of behavioral skills, and a public commitment procedure. Wolchik et al. (1993) describe a program for custodial mothers of children of divorce involving five components: the custodial parent–child relationship, the noncustodial parent-child relationship, discipline strategies, reduction of stressful events, and support from nonparental adults. Each of these programs has been evaluated in randomized trials comparing the intervention with a control group. Each program has shown some degree of success in producing its desired final outcome of reducing the prevalence of adolescent smoking and decreasing the children's symptoms, respectively. However, continued demonstrations of success will give rise to a new set of questions related to the multi-component nature of these programs.

THREE QUESTIONS ABOUT MULTI-COMPONENT PROGRAMS

The process of the development and evaluation of multi-component programs raises questions for both basic and applied scientists. The source of these questions for basic scientists is their concern with maximizing the informativeness of the results of the randomized trial of the intervention for basic psychological theory; the source of these questions for applied scientists is their concern about maximizing the effectiveness of the program in producing the desired outcomes. These concerns result in three intertwined questions that may be raised about multi-component programs and that will be considered throughout this article.

1. *Is each of the individual components of the program contributing to the outcome?* Basic scientists seek to show that each component of the program is producing the desired outcome in the service of establishing the construct validity of the independent variables (Cook & Campbell, 1979; Higginbotham, West, & Forsyth, 1988). Applied scientists may be interested in the effectiveness of individual components for more pragmatic reasons. Some components may have been included that in fact reduced the effectiveness of the overall intervention package in producing the desired outcome. Or, inert components may have been included that are neither

harmful or helpful, but that are costly to include in the program package. In such cases, applied scientists would desire to identify and delete such ineffective components from the overall program.

2. *Is the program optimal?* Program developers may wish to combine several promising components to develop a new program, to add promising components to an existing program, or to fine tune a successful program to achieve maximal effectiveness. Or, a program developer may have to operate within the constraints of a fixed amount of program time or program budget and wonder about how to allocate these resources to each of the components of the overall program. In each case, the program developer is seeking to identify the combination of components that produces the optimal outcome.

3. *Through what processes are the components of the program achieving their effectiveness?* Basic scientists are interested in understanding the processes through which each component may be achieving its effect on the outcome of interest. A new generation of prevention programs is explicitly being created based on psychosocial theory and research on the development and maintenance of the targeted problem (Caplan, Vinokur, Price, & Van Ryn, 1989; Sandler et al., 1992). Careful study of the processes through which preventive interventions achieve their effects potentially provides the strongest information possible to inform basic researchers about the development of psychopathology in children and adults (Coie et al., 1993). In addition, information about the processes through which the program operates may be critical in making appropriate modifications that help make the program successful in new sites.

The purpose of this article is to consider the strengths, weaknesses, and areas of application of a variety of designs that have been proposed to answer these three questions about multi-component programs. We begin by considering the often neglected background issue of statistical power, an issue that can place serious limits on the range of intervention designs that can be realistically considered. We then review traditional intervention designs discussed in the psychotherapy research literature. These designs have been adapted and used in the majority of randomized prevention trials reported in current psychological literature and can be considered to represent current practice. Turning to the statistics literature, we show that these traditional psychotherapy designs can be considered to be special cases of factorial and fractional factorial designs. Insights from the statistics literature are used to refine our understanding of what we can learn from the traditional psychotherapy designs and to address question 1 and question 2 in more depth. We then consider response surface designs that may suggest more sophisticated methods of addressing question 2 (program optimality). Finally, we consider strategies of examining mediation that combine experimental design and correlational approaches, addressing question 3 by providing an understanding of the process through which each component contributes to the outcome.

SOME IMPORTANT BACKGROUND: STATISTICAL POWER

For a randomized trial to be worth doing, it must have adequate statistical power to detect differences among intervention conditions. Following Cohen (1988), norms have been developed in the social sciences defining small, moderate, and large effect sizes as corresponding to a difference between treatment and control groups means of .20, .50, and .80 standard deviation units, respectively. A .80 or higher probability of detecting a specified effect size at $\alpha = .05$ is typically defined as adequate statistical power. Rossi (1990) reviewed articles in the 1982 volume of the *Journal of Consulting and Clinical Psychology* and found that mean power to detect small, moderate, and large effects was .17, .57, and .83, respectively. Other reviewers (e.g., Sedlmeier & Gigerenzer, 1989; West, Newsom, & Fenaughty, 1992) have reached similar conclusions about the power of statistical tests in other areas of psychological research. The implication of these results is that, with the exception of large effect sizes, the probability of detecting true differences between treatment conditions is virtual coin flip or worse in the typical study in psychology.

Many researchers continue to be unaware of the number of participants required to detect differences between treatment conditions with adequate power (see Aiken, West, Sechrest, & Reno, 1990). For example, in a randomized trial comparing an intervention and control group in which there are an equal number of participants in each condition, 52 total participants ($n = 26$ per cell) would be needed to detect a large effect, 126 participants would be needed to detect a moderate effect, and 786 participants would be required to detect a small effect on the outcome measure with .80 power and $\alpha = .05$.[2] User friendly statistical software (e.g., Borenstein & Cohen, 1988, Woodward, Bonett, & Brecht, 1990) is now available to provide *a priori* estimates of statistical power for commonly used intervention designs.

The work on statistical power holds several intriguing implications for the design of studies of the individual components of interventions. First, moderate to large sample sizes will be required to detect the moderate or

[2]These sample size requirements can be lowered by design improvements. The inclusion of a pretest measure that has a .5 correlation with the outcome measure in the above example lowers the total number of participants required to 19, 61, and 584 to detect large, moderate and small effect sizes, respectively, with .80 power.

small effects that apparently characterize many current preventive interventions (Durlak, Wells, Cotten, & Lampmann, 1993). These sample size requirements have little practical effect on large scale school-based or community-based primary prevention programs[3]; however, such requirements may restrict the complexity of the designs that may be contemplated for preventive interventions addressing more limited populations, notably populations with identified risk factors (e.g., bereaved children; Sandler et al., 1992). Second, initial comparisons of a full, multi-component program with a no treatment control will nearly always yield larger effect sizes than comparisons involving the effect of the inclusion of a single component over and above the effect of other components in an intervention package. Larger sample sizes will be needed in these latter designs to detect small effects. Third, developing efficient designs with contrasts focused on detecting the theoretically most important effects will have more power than omnibus comparisons of several treatment conditions to detect differences when, in fact, they do exist. To the extent such contrasts can be constructed to have equal sample sizes, their statistical power will be further enhanced.

CURRENT PRACTICE: TRADITIONAL PSYCHOTHERAPY RESEARCH DESIGNS

Kazdin (1980, 1986) has reviewed traditional intervention designs from the psychotherapy research literature. Below, we identify and review several approaches which have been adapted and used in nearly all randomized preventive trials. The first two of the designs are relatively common, whereas the latter three are currently only infrequently used in published prevention trials. An example of each approach is provided from the prevention literature where possible and from the clinical treatment literature when no instances of the approach could be located.

1. *Treatment Package Strategy.* In this approach, the effectiveness of the total treatment package is contrasted with that of an appropriate comparison group. For example, Wolchik et al. (1993) randomly assigned custodial mothers sampled from county divorce records to receive either the full intervention program or a delayed intervention (control group) begun after posttest data were collected. The full intervention package consisted

[3]Issues of the proper unit of analysis (Higginbotham et al., 1988; Shadish, 1992) may be raised for many large scale studies since they involve the assignment of units such as classrooms, schools, or communities to intervention conditions. These issues can be addressed through the use of hierarchical linear models (Bryk & Raudenbush, 1992; Kreft, 1992) if outcome data are collected from individual participants. These models adjust for the amount of dependency among cases within each unit, giving proper estimates of treatment effects.

of 13 group and individual sessions containing components that addressed each of the areas identified above (see p. 572). Such designs are ideal for determining whether the program works and is worthy of further research. Indeed, Sechrest, West, Phillips, Redner, and Yeaton (1979) have argued strongly for the use of this strategy to test initially what program developers believe is the strongest possible version of the program. However, this design by itself provides little information about the effectiveness of individual treatment components or the processes through which they operate.

2. *Comparative Treatment Strategy.* In the comparative treatment strategy, two or more alternative interventions are directly compared. An additional no treatment comparison group is often included in the design to enhance the interpretability of the results (Kazdin, 1986). The goal of this strategy is to choose the most effective single intervention from the set of alternative interventions under consideration. For example, Hansen, Johnson, Flay, Graham, and Sobel (1988) randomly assigned 84 school classrooms to receive one of three intervention conditions: (a) the social influences drug abuse prevention program (see p. 572), (b) an affective education program emphasizing stress management, values clarification, decision making, goal setting, and self-esteem building, or (c) no intervention (control). Such comparative designs can identify the most efficacious of a set of interventions as they were implemented in a particular randomized trial.[4]

Although Hansen et al. compared entire programs, the comparative design can also be applied to compare the effectiveness of potential components of a larger intervention package. When the design is used in this latter manner, it provides information about the unique effectiveness of each separate individual component. Such information can be useful to program developers in the design of a multi-component intervention package.

3. *Dismantling Strategy.* In the dismantling strategy (also termed the subtraction design) the full version of the program is compared with a reduced version in which one or more components have been eliminated. Criteria for selecting the component(s) to be deleted from the treatment package vary; however, they are often based on theory or other empirical work suggesting that the deleted component(s) may be inert or reduce the effectiveness of the retained components. Component(s) that are expensive or very difficult to deliver may also become candidates for deletion. Dis-

[4]More general interpretations about the relative efficacy of the interventions depend on meeting several important assumptions: The interventions should be of equal strength relative to the ideal treatment of that type, be implemented with equal fidelity, and should be expected to affect the same outcome variables (Cooper & Richardson, 1986; Sechrest et al., 1979).

mantling designs often add a third no treatment comparison group to enhance interpretability.

Pentz et al. (1989) provide an illustration of the use of this design to study the effectiveness of combinations of entire intervention packages. They designed a comprehensive community drug abuse prevention program from four component programs: (a) a school-based social influences program as one component (see p. 572), (b) a component training parents in positive parent–child communication skills, (c) a component training community leaders in the organization of a community drug prevention task force, and (d) mass media coverage. The comprehensive program was compared with a reduced, lower cost version that only included components (c) and (d). To the extent that the reduced version of the program produced outcomes that did not differ from the comprehensive program, but did differ from a no treatment comparison group, the researchers would be justified in concluding that the addition of programs components (a) and (b) did not add to the effectiveness of the comprehensive program *over and above* that of the reduced program comprised of components (c) and (d).[5]

4. *Constructive Research Strategy.* In the constructive research strategy, one or more components are added to a base intervention. The base intervention may be a single component or it may be an entire program package. Added components that increase the effectiveness of the base intervention are retained by program developers, whereas those that do not improve or which decrease effectiveness relative to the base intervention are discarded.

To illustrate, Perri et al. (1988) examined the effects of several components designed to help maintain weight loss in obese adults. All participants received the base intervention, a 20-week behavior therapy program [B]. Participants were then randomly assigned to receive one of four combinations of weight loss maintenance components or a fifth, control condition consisting of no additional maintenance components. The four maintenance interventions examined were (a) bi-weekly therapist contact [C], (b) bi-weekly therapist contact plus a social influence component [S] designed to enhance the participant's motivation, (c) bi-weekly therapist contact plus aerobic exercise component [A], and (d) bi-weekly therapist contact plus social influence component plus aerobic exercise component.

[5]As will be discussed in section below on factorial and fractional factorial designs, this comparison is not informative about the effectiveness of components A, B, or A + B considered alone. The effect of the full program reflects the main effect of each component taken separately plus all possible interactions among the components. The effect of the reduced version of the program only reflects the main effects and interactions among the components that are present in the reduced program.

Thus, the five conditions of this constructive research study can be described as B, BC, BCS, BCA, BCAS.

Like the dismantling strategy, the constructive strategy can provide information about the effectiveness of adding individual intervention components over and above a base intervention. Indeed, in the minimal versions of these designs that are presently represented in the literature (e.g., comparing an intervention comprised of component A with one comprised of components A and B), the designs can be distinguished only by whether the researchers take A (constructive) or A + B (dismantling) as the base comparison group. The information about the effectiveness of individual components provided by the constructive strategy depends on the theoretical rationale for the selection of components, the number of components that are added in each comparison, and the particular combinations of components that are selected relative to the full set of possible combinations.

5. *Factorial Designs.* Complete factorial designs have long been among the most commonly used designs in laboratory experiments in psychology; they have also attracted modest attention in the psychotherapy research literature (Kazdin, 1980). In these designs, interventions representing all possible combinations of the levels of one component (factor A) and the levels of a second component (factor B) are created. For example, Webster-Stratton, Kolpacoff, and Hollinsworth (1988) randomly assigned parents of conduct problem children to one of four conditions in a 2×2 factorial design: (a) videotape modeling of parenting skills plus group discussion, (b) videotape modeling only, (c) group discussion only, and (d) a waiting list control group. This design permits separate estimates of the effects of the videotape modeling component (factor A), the group discussion component (factor B), and their interaction.

Factorial designs potentially represent a powerful approach to the examination of the separate and combined effects of treatment components. However, virtually all of the factorial designs in the published literature on intervention trials are limited to 2×2 designs involving the presence vs. absence of two intervention components. The use of factorial designs in the investigation of more complex multi-component interventions would appear to be a natural extension of such previous research. However, as will be discussed in the next section, the complexity of the resulting designs, the difficulty in mounting the large number of treatment combinations, and the large number of participants required for adequate statistical power have thus far limited the use of full factorial designs.

INSIGHTS FROM THE EXPERIMENTAL DESIGN LITERATURE

The review of traditional intervention research designs has provided several insights about how the first two questions posed in the introduction can be addressed. Additional insights can be gleaned from a consideration of new and recycled ideas from the experimental design literature in statistics and in psychology (Box & Draper, 1987; Box, Hunter, & Hunter, 1978; Mead, 1988; Myers, Khuri, & Carter, 1989; Pilz, 1983; Steinberg & Hunter, 1984; Woodward, Bonett, & Brecht, 1990). Two areas are of particular interest. First, we initially limit our consideration to designs in which components can only be included or not included in an intervention package. For these designs the work on factorial and fractional factorial designs allows us to extend and refine the insights from the traditional psychotherapy research literature in answering the first two questions outlined in the introduction. Second, we consider the possibility that the strength of the intervention components can be quantitatively scaled so a range of strengths of each component can be considered. For this case ideas from work on response surface designs that may provide particularly strong answers to question 2.

Factorial and Fractional Factorial Designs

We have previously noted the problem that complete factorial designs in which each component is separately manipulated rapidly become too complex to implement. To illustrate this problem, consider developing a factorial design for the Wolchik et al. (in press) custodial parent-based program for children of divorce which involved five components (see p. 572). Each of the five components would be independently manipulated to be present or absent in the treatment condition. This strategy gives rise to a 2^5 ($2 \times 2 \times 2 \times 2 \times 2$) factorial design with 32 treatment conditions. The design would be analyzed with analysis of variance (ANOVA) giving rise to five main effects (corresponding to the main effect of each separate component), ten two-way interactions, ten three-way interactions, five four-way interactions, and one five-way interaction. Such a design would typically not be practical because of both the difficulty in implementing the large number of intervention conditions and the very large number of participants that would be required to achieve adequate statistical power for the tests of the interactions. Further, the design is likely to be inefficient since researchers almost never have theoretical or empirical expectations that the higher order (three-way, four-way, and five-way) interactions will be significant.

Considerable work in the statistics literature indicates that these complex full factorial designs may be simplified to fractional factorial designs in which only a systematically selected portion of all possible treatment combinations are implemented. Such simplification requires that the researcher be willing to assume that certain effects, typically higher order interactions, are negligible. Indeed, all of the traditional intervention designs from the psychotherapy research literature discussed above can be considered to be special cases of fractional factional or full factorial designs. This fact helps clarify the assumptions underlying the traditional designs as well as providing a basis for suggesting improvements to the traditional designs.

We illustrate the use of three of these simplified fractional factorial designs to investigate three of Wolchik et al.'s (1993) five intervention components: (a) custodial parent–child relationship, (b) discipline strategies, and (c) stressful events. In Table I, we show the combinations of conditions that constitute the design; these combinations are always a subset of the eight (2^3) unique combinations of the complete factorial design that could be created from the three intervention components. In keeping with our focus in this section on designs in which each program component can only be either present or absent in the intervention package, we designate those components that are present in the package with "yes" and those that are excluded with "no." For example, A = yes, B = no, and C = yes means that the custodial parent–child relationship and the stressful events, but not the discipline strategies components were included in the intervention package.

The first example of a fractional factorial design is illustrated in Table I(A). This design addresses question 1, comparing a set of intervention conditions, each comprised of a different single component, with a no treatment comparison group. Note that we have reduced the full factorial design to a comparative treatment design (see p. 576) that contrasts components A, B, and C with a control group. This design provides unbiased estimates of the effect of each component separately, but *only in the absence of any of the other components.* Without making assumptions that all two-way and three-way interactions are negligible, predictions cannot be made about the effectiveness of combinations of treatment components.

A second fractional factorial design is illustrated in Table I(B). This design adds each component sequentially to the treatment package. A no treatment control group is compared with groups receiving only component A, components A and B, and components A, B, and C. This design is identical to the most commonly used version of the constructive strategy (see p. 577) with three components. Recall in this design, each test reflects the contribution of the new component *over and above* the components that are already included. For example, the comparison of the A + B + C intervention with the A + B intervention tests the effectiveness of what

Table I.

A. Comparative treatment strategy including no treatment control group

Condition	A	B	C
1	no	no	no
2	yes	no	no
3	no	yes	no
4	no	no	yes

B. Constructive research strategy including no treatment control group

Condition	A	B	C
1	no	no	no
2	yes	no	no
3	yes	yes	no
4	yes	yes	yes

C. Two fractional factorial designs that permit main effect estimates

	Block 1				Block 2		
Condition	A	B	C	Condition	A	B	C
1	no	no	no	5	no	no	yes
2	no	yes	yes	6	no	yes	no
3	yes	no	yes	7	yes	no	no
4	yes	yes	no	8	yes	yes	yes

component C adds, given that A and B are already present in the treatment package. This design does not provide tests of the unique effect of each component unless it is assumed that the three two-way and one three-way interactions among the components are negligible.

A third fractional factorial design illustrated in Table I(C) is unfamiliar to most intervention researchers in psychology. Two different examples (Block 1; Block 2) of this type of design known as the 2^3–1 design (half fraction; Box, Hunter, & Hunter, 1978) are presented in the left and right halves of Part C of the table. This design provides unbiased estimates of main effects of each treatment component if we assume that all interactions are negligible. Given this assumption, the four conditions in *either* Block 1 or Block 2 provide unbiased estimates of the main effects of each of the three components.

Readers should recognize that the three fractional factorial designs described above answer different questions and have different strengths and weaknesses. Design A provides an answer to question 1 in that it informs us about the unique effects of each component; however, it provides no information about the effectiveness of intervention packages comprised of combinations of the components unless it is assumed that the components do not interact. This deficiency can be partially remedied in the general case by adding a fifth condition to the design, A = yes, B = yes, C = yes. This added condition provides information about whether the sum of the three two-way and one three-way interaction effects is 0.

Design B provides a partial answer to question 2. This design informs us about one specific sequence of building up the intervention components; however, it does not provide information about the effects of individual components unless it is assumed that they do not interact. Further, other non-examined intervention packages (e.g., A = no; B = yes; C = yes) could potentially be even more effective than any of the set of interventions that were examined. There is no guarantee that the set of tested combinations of components will include the program representing the optimal combination of components.

Design C offers a higher level of statistical power in testing question 1 than does Design A. However, its interpretation requires the strong assumption that the components do not interact. An interesting feature of this design is that the two versions illustrated are complementary. If Block 1 of Design C were used in the initial intervention trial and Block 2 were used in a replication, the full factorial design is constituted across the two studies in an economical manner. The Block (replication) effect in this case is unconfounded with all main effects and two-way interactions. If there were a main effect of Block stemming, for example, from the use of a healthier population in study 2 than study 1, the only effect estimate that would be biased is the three-way interaction.

More generally, fractional factorial designs can be constructed to permit economical tests of a specific set of effects of interest given the assumption that all other effects are negligible. For example, if a researcher were interested in testing the main effects of components A, B, and C and the A × B interaction, the five condition design illustrated in Table II provides unbiased tests of each of these effects. Box, Hunter, and Hunter (1978) describe general methods for constructing fractional factorial designs that provide unconfounded estimates of main effects and two-way interactions if higher order interactions are assumed to be zero; Anderson and McLean (1984) provide a cookbook of these designs. These sources should be consulted to develop customized designs that permit tests of the specific effects of interest to the investigator. Many of the designs are very economical relative to the full factorial designs. For example, consider the social influences smoking prevention programs described earlier which has six components. Box et al. (1978) describe a 16 cell design (quarter fraction) that provides unbiased estimates of all main effects all and two-way interactions for six factors, each having two levels, assuming all three-way and

Table II. Fractional Factorial Design: Estimates A, B, and C Main Effects and A × B Interaction

Condition	A	B	C
1	no	no	no
2	no	yes	no
3	yes	no	no
4	yes	yes	no
5	no	no	yes

above interactions are negligible. This design is distinctly more feasible than the full 2^6 factorial design which requires 64 cells. At the same time, prevention researchers will rarely be able to mount large enough trials to permit even 16 intervention combinations to be investigated. Even these reduced designs can become impractical if several main effects and two-way interactions are of interest.

Comment. Each of the fractional factorial designs adequately addresses a version of either question 1 or question 2. However, since these are not complete factorial designs, additional assumptions must be made in each case to answer more general questions, particularly those involving component combinations not included in the design. If these assumptions are not reasonable, the estimates of effects of interest will be biased because they will be confounded by higher order interactions. This same issue applies to the first four traditional intervention designs reviewed in the previous section. Researchers need to be attentive to the possibility that nonzero interactions among components have the potential to alter their conclusions.

The examples of factorial and fractional factorial designs discussed in this section have all involved only two levels of each component. This design decision implicitly makes the assumption that each of the treatment components can only have linear effects on the outcome variable (Aiken & West, 1991). Fractional factorial designs can be extended to address factors having more than two levels (Anderson & McLean, 1984; Box, Hunter, & Hunter, 1978), again at a cost of requiring a large number of intervention combinations.

Dose Response. Response Surface, and Optimal Designs

Another class of potentially useful design approaches is applicable for researchers interested in addressing question 2, if each component can be scaled on a *continuum* of strength relative to the ideal version of the component. This strategy is often followed in drug research in which the outcomes produced by conditions representing three (or more) levels of dosage of the drug (typically including a no dosage placebo control) are compared. For example, Whalen et al. (1987) compared the social behaviors of hyperactive children who had received a placebo, a low dose, or a high dose of methylphenidate. Such designs can identify components that have a optimal strength beyond which further increases in strength lead to either decreases or no further increases in effectiveness.

Sechrest et al. (1979) have proposed that many types of interventions could also be scaled on a dimension of treatment strength and have suggested methods for doing this. To cite a straightforward example, Shure (1988) has argued that a session of her school-based interpersonal cognitive problem solving (ICPS) intervention for young children should be delivered daily for approximately 12 weeks duration to achieve maximum effectiveness. Other researchers have used markedly shorter versions of the program with far less impressive outcomes. An experiment could be designed in which the duration of the program was varied, leading to a dose-response curve relating program duration to outcome. The selection of the levels of the strength of the treatments that are compared would depend on theory or prior empirical work. For example, if researchers expected the dose response curve to be linear and wanted to run four conditions, they might use 0 (no treatment control), 4 weeks, 8 weeks, and 12 weeks. On the other hand, if they had a strong expectation that the curve would be of an exponential form in which the outcome initially increases very quickly followed by a tapering off of the rate of increase with additional sessions, they might consider using 0 weeks, 1 week, 2 weeks, 4 weeks, and 12 weeks.[6] Of course, other cost-related or practical criteria (e.g., what length program is the school willing to consider?) could be used in the design of the dose–response intervention trial.

Dose response experiments can also be generalized to more than one dimension of treatment. This generalization requires that each of the intervention components be scaled on a continuum of treatment strength. If we construct several interventions representing a number of combinations of different levels of strength of each treatment component and plot the outcome for each combination, the resulting figure would be known as a response surface. To illustrate the response surface design, imagine that the school-based component now represents only the first component of

[6]If the program content is generally uniform, repeated measures designs provide more efficient estimates of the dose response curve. However, if the program involves several sequential phases with the duration of the program being determined by the amount of time spent on each phase (e.g., amount of practice), between subjects designs may be preferable (see Greenwald, 1976). Response surface methods applied to programs with generally uniform content are probably best implemented as mixed between-within designs.

the ICPS program, which may be administered for varying lengths of time up to 30 weeks. The second component consists of a home-based program in which parents also train their children in ICPS skills for up to 12 weeks. The response surface representing the relation between each possible combination of durations of (a) the school-based component, and (b) the home-based component to the level of the child's outcome on a measure of adjustment can be plotted. Figure 1 represents several *hypothetical* response surfaces. In Fig. 1a, the effectiveness of the home-based program increases linearly with the duration of the program, whereas the school-based program has no effect. In Fig. 1b the duration of both the school-based and home-based programs are linearly and additively related to child adjustment. In Fig. 1c both programs have positive effects when each is presented separately, but they have a negative interaction when combined such that the combined effect substantially is less than the effects of either component when delivered separately. Finally, Fig. 1d represents curvilinear (quadratic) effects such that the maximum effectiveness is obtained for an 18-week school-based component combined with an 8-week home-based component. These hypothetical response surfaces clearly illustrate a few of the complexities that potentially may arise when treatment components are combined in a program.

Response surface methodology (Box & Draper, 1987; Myers, Khuri, & Carter, 1989) provides an approach to identifying which combination of components produces the optimum outcome. This approach is also useful for studying the tradeoffs in outcomes that occur when the resources allotted to each program component are varied within an intervention package in which the total duration or the total costs of the program are fixed. As is illustrated by Fig. 1c, conducting separate dose response experiments on each component often may not identify the intervention package that includes the combination of components that produces the maximum outcome. Box, Hunter, and Hunter (1978, pp. 510-513) present an extensive illustration of the limitations of individual dose response experiments in the identification of the optimal combination of components.

Returning to our example, assume that the duration of the school-based (component A) and home-based (component B) components may have linear or quadratic relations to adjustment. Further, the two components may interact so long as the form of the interaction is linear by linear. Under these constraints, the researchers might use the four corners plus central point design illustrated in Table III, which can probe each of these effects. Note that the levels of each of the duration variables listed in Table III are simply called low, medium, and high. If a prior empirical or theoretical basis existed for describing the response surface, then optimal values for the durations of each component in each of the five conditions to maxi-

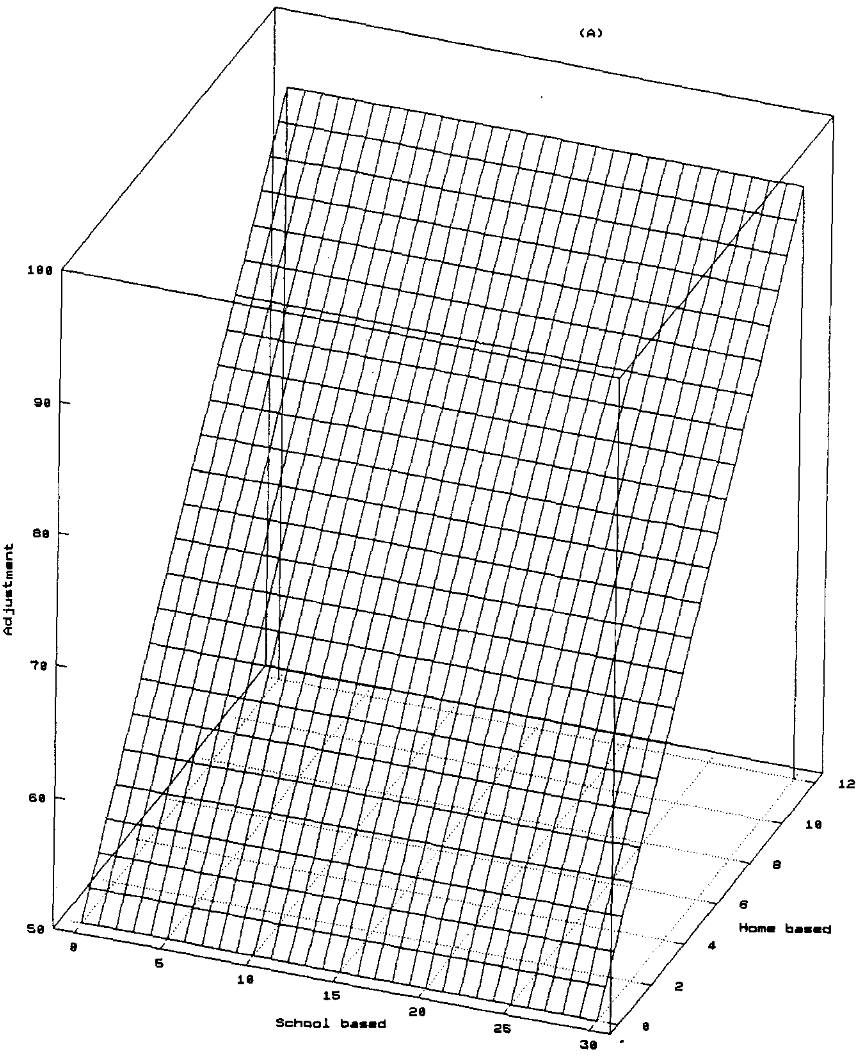
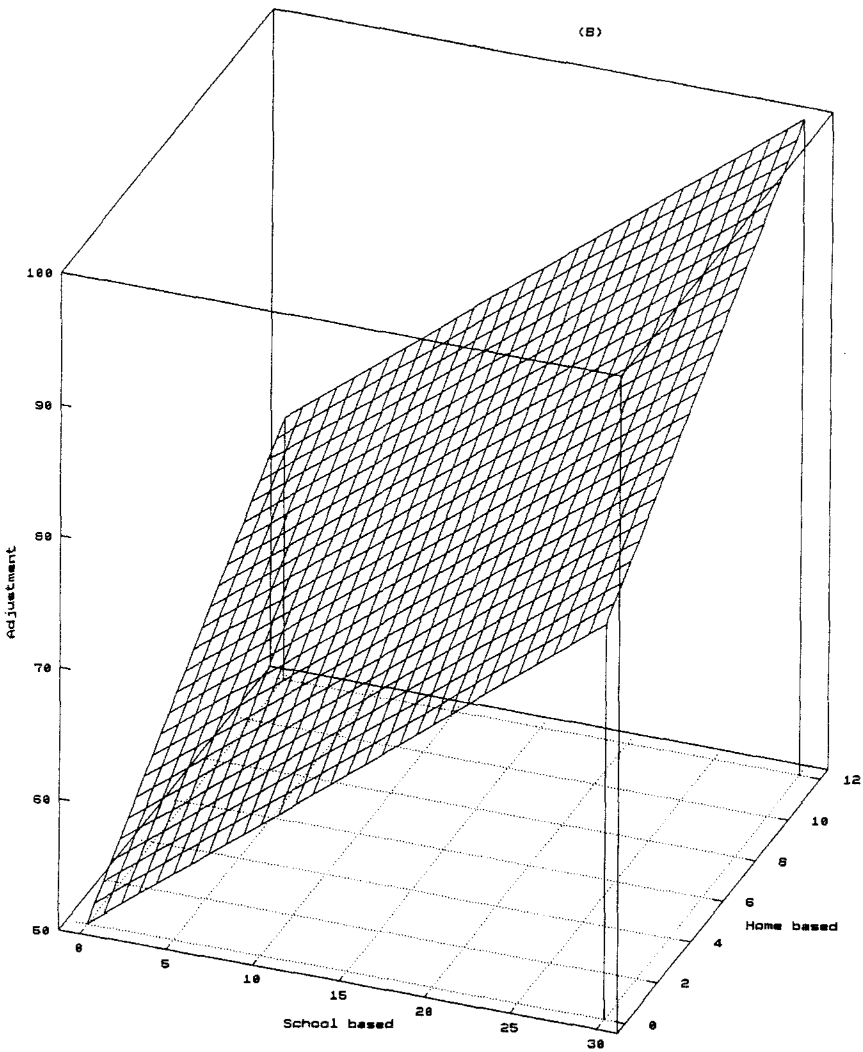
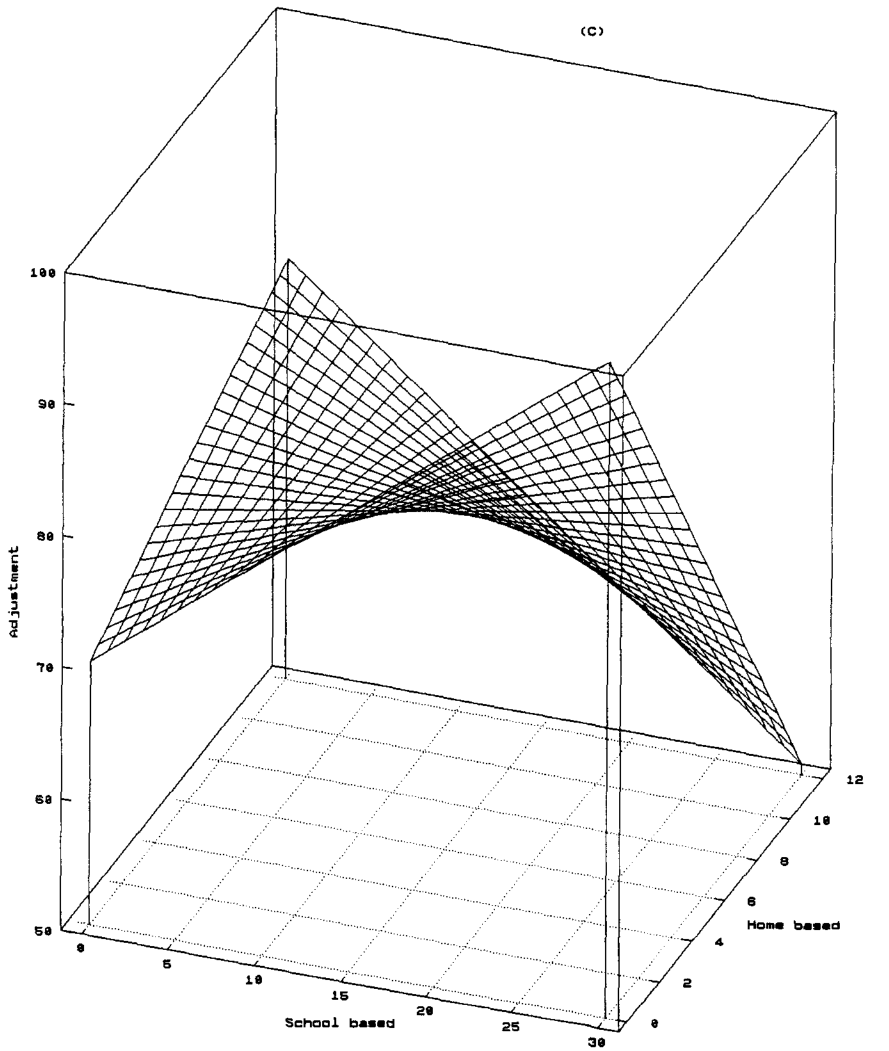


Fig. 1. Four hypothetical response surfaces. Each hypothetical response surface represents the level of adjustment of children receiving various combinations of the school-based (0–30 weeks) and the home-based (0-12 weeks) program. (A) represents a linear effect of only the home based program. (B) represents linear effects of both the school-based and home-based programs. (C) represents a negative interaction between the school based on home-based programs. (D) represents curvilinear effects of both the school-based and home-based programs.

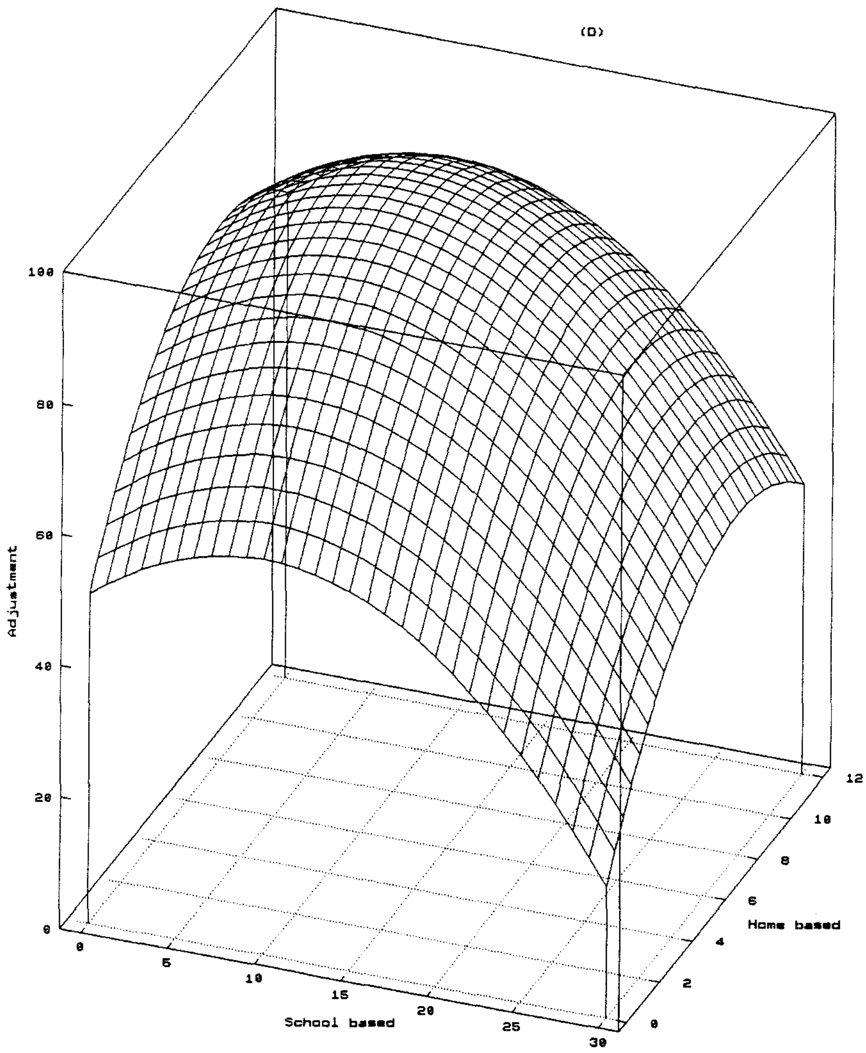


mize the power of the test of each potential effect can be statistically specified (see Atkinson, 1985; Pilz, 1983; Silvey, 1980). In the absence of prior knowledge, the five conditions could be placed at the extreme values for each component (A = 0, B = 0; A = 0, B = 12; A = 30, B = 0; A = 30; B = 12) and the approximate midpoint of the two components (A = 15; B = 6). The response surface can then be plotted, permitting the researcher to provide a preliminary estimate of the combination of levels of



components under which the maximum (or minimum) response should occur.[7]

[7]Researchers studying a positive outcome such as self-esteem or social competence would have interest in the maximum point, whereas researchers attempting to decrease a negative outcome such as symptoms would search for the minimum point. For ease of presentation, we assume the researcher is interested only in the maximum point in our discussion.



An interesting feature of response surface methodology as it has been applied in engineering and applied biology is that the initial answer is considered to be only preliminary (Mason, Gunst, & Hess, 1987). The statistical theory prescribes methods for designing a sequence of experiments to help pinpoint the combination of levels of the components that produce the maximum point on the response surface (optimum intervention). However, given the current stage of development of prevention research, the use of these sophisticated statistical procedures is premature. But, three lessons from response surface methodology are potentially very important. (a) Increasing the strength of a component does not always lead to corresponding increases in effectiveness. Similarly, combining two individually effective components may lead to a program that is either more or less effective than the individual components. (b) Programs can be improved sequentially through refinement of each of the individual components and the study of their combined effects. Development of an optimal program is an evolutionary process. (c) In addition to studying programs whose effectiveness is at the apparent maximum of the response surface, it is also sometimes useful to contemplate interventions representing areas of the response surface that have not actually been studied. If there is a strong theoretical rationale for a specific combination of components and the plotted theoretical response surface appears to be increasing around the particular combination of interest, then there is a reasonable chance that the new combination of components may produce optimal or near optimal effectiveness.

Comment. Response surface methodology is the best method for identifying the optimal combination of a set of treatment components which have been quantitatively scaled. Difficulties may arise in attempting to quantitatively scale many intervention dimensions. In addition, interventions often have multiple outcomes which may vary in their degree of positivity. That is, different packages of intervention components may produce the optimal result on each of several different outcome measures. Or, the outcomes from different treatment packages may differ over time with intervention A showing the best result at immediate posttest, whereas intervention B produces the best result at 1-year followup. Such issues can be addressed by selecting the single most important criterion of outcome, by defining a single aggregate outcome measure, or by choosing a set of weights to represent the importance of each outcome measure (see also Myers et al., 1989). Although the most sophisticated applications of re-

Table III. Four Corners Plus Center Design[a]

		A. Duration of cognitive problem solving		
		Low	Moderate	High
B.Duration of parent training	Low	Yes	—	yes
	Moderate	—	yes	—
	High	yes	—	yes

[a]Yes = treatment condition in included in design;—treatment condition is omitted from design.

sponse surface methodology are beyond the current stage of development of prevention research, many of the concepts from this methodology provide presently useful design insights. For example, the Four Corners plus Center design shown in Table III offers a good initial picture of a variety of response surfaces including all those depicted in Fig. 1. The concept of the evolution of designs based on prior theory and data permits us to fine tune a program to achieve optimal effectiveness. These basic concepts underlying response surface methodology have occasionally been used in prevention research with considerable success. For example, Tharp and Gallimore (1979; see also Fienberg, Singer, & Tanur, 1985 for statistical commentary) describe a successful 10-year project in which they developed an early educational intervention program for native Hawaiian school children based on adaptations of several of the fundamental concepts underlying response surface methodology.

MEDIATIONAL ANALYSIS

Mediational analysis provides an economical, but less definitive approach to question 1 than factorial and fractional factorial designs. It also provides an excellent method of addressing question 3, the process through which each component has its influence.

In mediational analysis, the researcher articulates what Lipsey (1992; see also Wolchik et al., 1993) terms a "small theory" that specifies the processes that are targeted by each component of intervention. The relation of each of these processes to the outcome(s) of interest is then specified. Reliable measures of each of the processes (putative mediators) and each of the outcome variables are included in the design. Through statistical techniques such as structural equation analyses the researcher has some ability to probe the contribution of each of the putative mediators to the outcome.

To understand mediational analysis, it is useful to consider initially the case of a simple, one component program. Imagine that a hypothetical training program is expected to improve children's social skills and that these improved social skills, in turn, are expected to reduce the children's level of aggressiveness. Figure 2a depicts this set of relations which represent the small theory of this simple program. The researchers conduct a randomized trial comparing program participants with a no treatment control group. Each child's level of social skills and aggressiveness are assessed after completion of the program.

Following Judd and Kenny (1981a), three conditions must be met to demonstrate that social skills mediated the outcome.

1. The program must cause differences in the putative mediator, here the measure of social skills. This can be tested with a simple two group Analysis of Variance (ANOVA) or equivalently by regression analysis with a binary predictor (intervention, no intervention), which, if significant, shows that the program affected the mediator. The standardized regression coefficient from such an analysis is the path coefficient *a* in Fig. 2a.

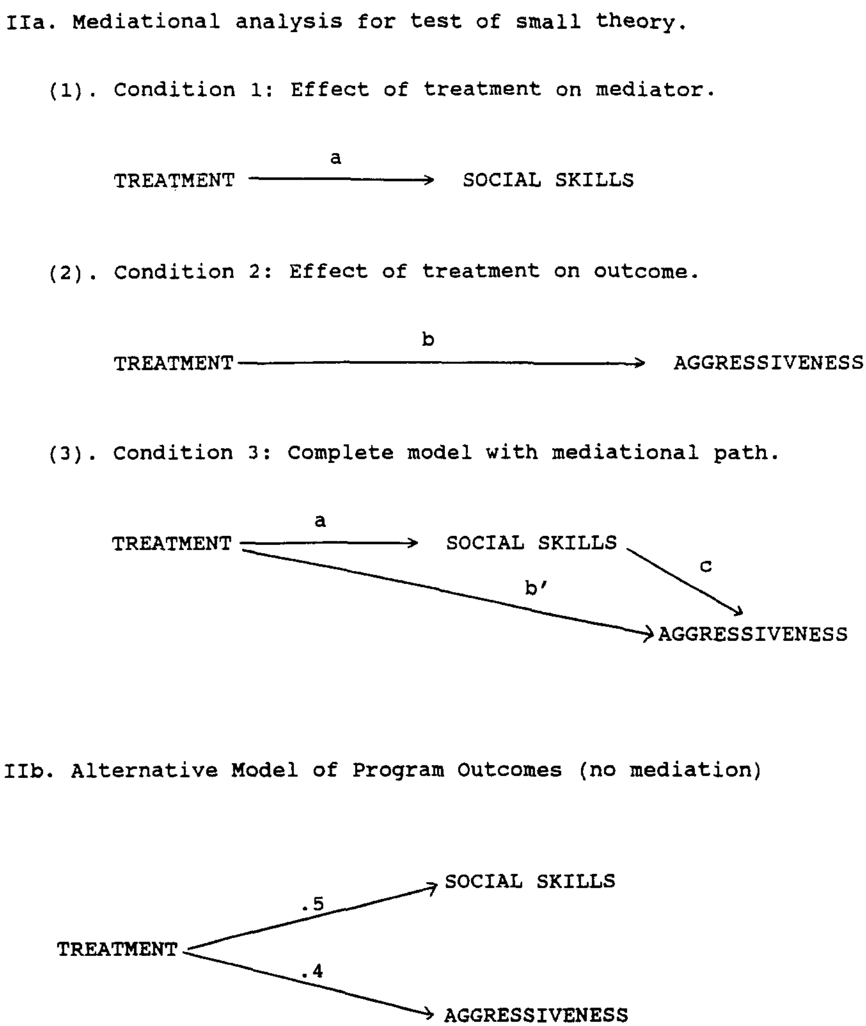


Fig. 2. Mediational analysis of small theory of a program.

2. The program must cause differences in the outcome, here aggressiveness. Again, this can be tested with ANOVA or with regression analysis, yielding the path coefficient *b* in Fig. 2b.

3. The links from intervention to social skills to aggressiveness (paths *a* and *c* in Fig. 2c) represent mediation. When these paths are controlled, the magnitude of path b' must be significantly reduced. If path b' does not differ from 0, then social skills may be inferred to be fully mediating the effect of the program on aggressiveness. If the magnitude of path b' is significantly reduced relative to its value (*b*) in the test of condition 2, social skills only partially mediate the effect of the program on aggression. This result suggests that at least one other mediator that affects the outcome is also affected by the training program. This other mediator may represent other unmeasured skills taught by the program or other unmeasured effects such as the child's relationship with the intervenor. This third condition can be tested using multiple regression or structural equation modeling (see Baron & Kenny, 1986).

Conditions 1 and 2 are straightforward and have long been tested by researchers with mediational hypotheses. A few researchers have also tested what may appear to be an alternative to condition 3, namely that the putative mediator must be correlated with the outcome variable. This condition, in fact, must hold true if mediation is taking place. However, results in which mediation is *not* taking place can also meet this alternative condition. To illustrate, consider the result depicted in Fig. 2d in which the program has two *independent* effects: an increase in social skills and a decrease in aggression. This result passes conditions 1 and 2. Further, since social skills and aggression share the common third variable of program status, they will be correlated: $r = .20$ in this example (see Duncan, 1975). Only by imposing condition 3 can we rule out this and some other possibilities that are not consistent with the small theory of the program.

Mediational analysis is illustrated in a study by Harackiewitz, Sansone, Blair, Epstein, and Manderlink (1987) who compared the effectiveness of four smoking cessation program packages: (a) nicotine gum plus a self-help manual with an intrinsic motivational orientation; (b) nicotine gum plus a self-help manual with an extrinsic motivational orientation; (c) intrinsic self-help manual only; (d) a brief booklet containing tips for stopping smoking (control). Measures of one putative mediator, attributions for success or failure in quitting smoking, were collected 6-weeks after intake; follow-up measures of smoking status were collected at regular intervals up to 1 year after intake. The results were consistent with Judd and Kenny's (1981a) three conditions and suggested that intrinsic attributions for success partially mediated successful maintenance of nonsmoking status.[8]

The extension of mediational analysis to multiple component intervention programs raises new issues, particularly ones associated with the simultaneous investigation of the effects of more than one putative mediator. The small theory of the intervention typically becomes considerably more complex. Exactly how to apportion variance among competing mediational paths becomes less definitive. The statistical tests of the model also increase in difficulty and the impact of problems in study design or measurement of the mediators becomes more serious.

To illustrate some of these issues, we consider partial data from a trial of the second generation of an educational program originally developed by Reynolds, West, and Aiken (1990) to increase the incidence of screening mammography. In this trial Aiken and West (1993) exposed eligible women to an intervention package that included components designed to influence the participants' *perceptions* of four putative mediators proposed by the Health Belief Model (HBM): susceptibility to breast cancer, severity of breast cancer, benefits of screening mammography, and barriers to screening mammography. Each of these putative mediators, in turn, was expected to influence the participants' intentions to get a screening mammogram. Eligible women (N = 135) were assigned to intervention or control groups, and their level on each of the four putative mediating variables and the outcome variable of intentions to get a screening mammogram were assessed immediately after the presentation of the program.

Applying the three conditions of Judd and Kenny (1981a) to these data, we observe the following results (see Fig. 3).

1. In the test of the relations of intervention to the mediators (see Fig. 3a), the intervention package led to higher perceptions of two of the putative mediators, susceptibility to breast cancer, $\beta = .27$, $F(1, 132) = 10.80$, $p < .005$, and benefits of screening mammography, $\beta = .27$, $F(1,132) = 10.32$, $p < .005$. Neither perceived severity of breast cancer, $\beta = .04$, nor perceived barriers to screening mammography, $\beta = -.10$, were significantly affected by the intervention.

2. The intervention package led to a significant increase in the outcome variable, intentions to get a screening mammogram, $\beta = .51$, $F(1, 132) = 43.99$, $p < .001$ (see Fig. 3b).

[8]Complications in this conclusion arise because of the focus of the analysis only on those participants who had successfully quit 6 weeks after intake and the nature of the outcome measures (smoker vs. nonsmoker; duration of nonsmoking status [time to failure]). The latter problem can be addressed through the use of alternative analysis strategies to test conditions 2 and 3 (see MacKinnon & Dwyer, 1993).

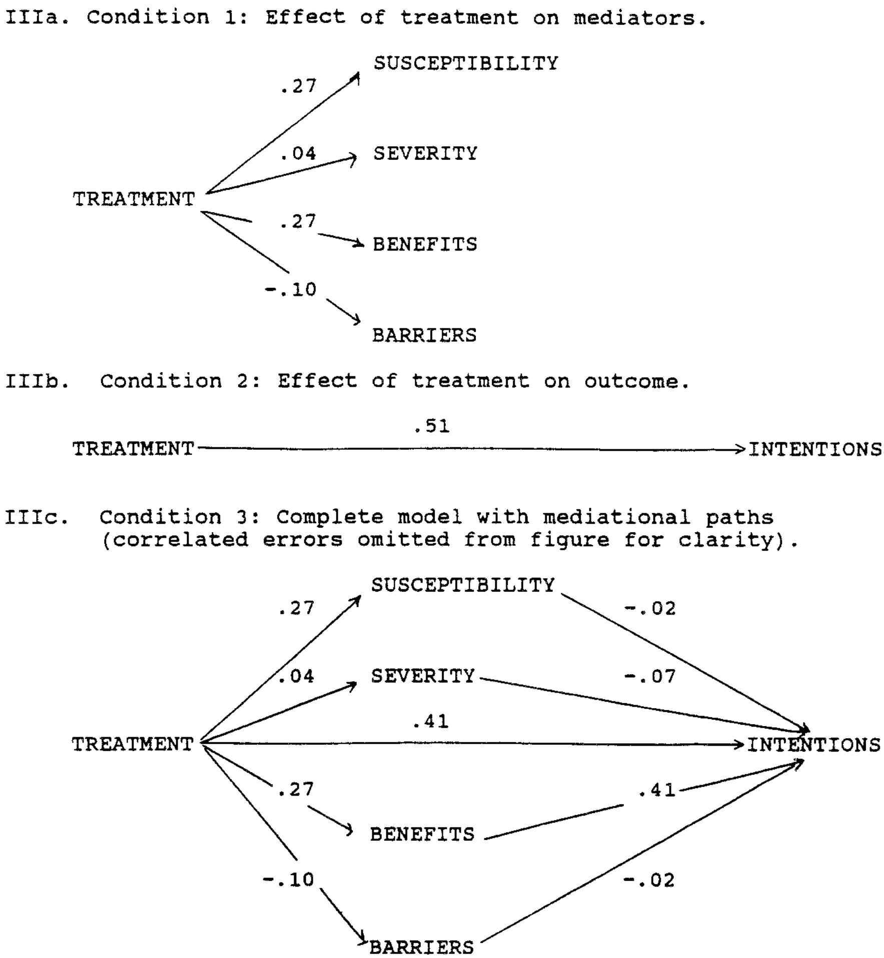


Fig. 3. Sequence of models used to examine mediation of outcomes in multicomponent program. Note: Correlated errors between putative mediators are omitted from figure.

Three of the putative mediators, perceived susceptibility to breast cancer, r = .18, perceived benefits of screening mammography, r = .52, and perceived barriers to screening mammography, r = –.23, showed significant relations with intentions to get a screening mammogram. Perceived severity of breast cancer was not significantly related to intentions, r = –.02, ns. The results of these initial analyses suggest that perceived susceptibility and

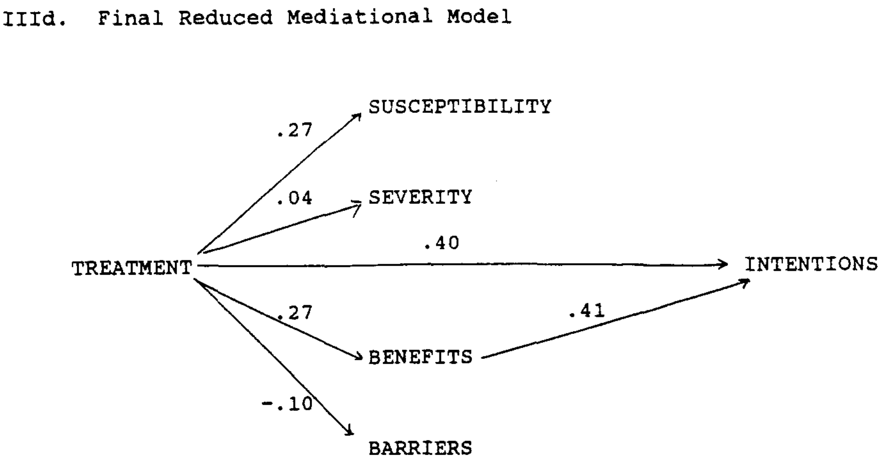


Fig. 3. Continued.

perceived benefits are candidate mediators of the effect of the intervention on intentions.

3. To test the third condition proposed by Judd and Kenny (1981a), we constructed the structural equation model depicted in Fig. 3c. In this model the intervention had indirect paths through each of the four putative mediators specified by our small theory as well as a direct (unmediated) path to intentions. In addition to the paths that are depicted, we allowed the errors of measurement between each pair of putative mediators to be correlated. The initial test of the just identified model indicated that *none* of the putative mediators had even a marginally significant ($p < .10$) path to intentions. However, note also that the effect of the intervention on intentions was also no longer significant, $\beta = .41$, ns. This result illustrates one complexity that can arise in probing the effects of multiple putative mediators: None of the putative mediators demonstrated a significant path to outcome, yet the effect of the intervention was substantially reduced suggesting that partial mediation may be taking place.

To understand these results further, we tested a reduced model that omitted the four paths from the mediators to the outcome. This model permits only a direct effect of the intervention on intentions. A χ^2 difference test (Bentler & Bonett, 1980) comparing the model depicted in Fig. 3c and this initial reduced model showed that the addition of the paths from the four putative mediators to the outcome improved the fit of the model, $\chi^2(4) = 14.50$, $p = .005$. This analysis suggests that mediation is occurring through the set of measured mediators, but does not identify which of the mediators are accounting for the effect.

To explore this question further, we tested two models in which the path from one of the two candidate mediators (perceived susceptibility; perceived benefits) identified above to the outcome was added to the model. In these models the intervention has its effect on intentions only through the direct path and the single indirect path through the putative mediator under consideration. The chi-square difference tests were $\chi^2(1) = 2.34$, ns for perceived susceptibility and $\chi^2(1) = 13.04, p < .001$, for perceived benefits. Of interest, the final reduced model including the indirect path through benefits to intentions depicted in Fig. 3d provided an adequate fit to the data, $\chi^2(3) = 1.46$, ns, CFI $= 1.00$. The path from perceived benefits to intentions was significant, $\beta = .41, p < .05$; and the path from the intervention to intentions to get a mammogram was marginal, $\beta = .40, p < .10$. These results are consistent with an interpretation of partial mediation in which benefits is the only putative mediator that meets all conditions for mediation. However, caution must be exercised in interpreting these results: Other models, for example, involving interactions between program components are *not* ruled out by these analyses.

The above example illustrates the extension of mediational analysis to the probing of multiple potential mediators between the intervention and the outcome variables. Mediational analysis can also be extended to test small theories that propose longer causal chains between the intervention and the outcome variable. For example, Judd and Kenny (1981b) outline a mediational analysis of the Stanford Heart Disease prevention project (Maccoby & Farquhar, 1975) in which (a) mass media and personal interventions were expected to (b) increase knowledge about diet and heart disease. This increased knowledge, in turn, was expected to change (c) participant's dietary behavior, which in the long run was expected to lead to (d) lowered physiological indicators of risk for heart disease (blood cholesterol and triglyceride levels). Although the logic of this extension is straightforward, to the authors' knowledge no examples of the actual mediational analysis of preventive trials expected to operate through longer causal chains are currently available in the literature (but see Aiken, West, Woodward, Reno, & Reynolds, 1993, for an example).

Comment. Mediational analysis provides an economical method for probing the contributions of individual components and the processes through which they operate to produce the outcome. Such analyses have now been used successfully for this purpose in several preventive intervention trials (Aiken et al., 1993; MacKinnon et al., 1991; Wolchik et al., 1993). At the same time, the limitations of this technique need to be clearly recognized.

Mediational analysis as applied in preventive trials is a hybrid between experimental and correlational (structural equation) approaches. Judd and Kenny's (1981a) conditions 1 and 2 are experimental tests whose interpretation is predicated only on the success of the random assignment to the intervention groups. The test of condition 3 is at its heart correlational; its interpretation depends on meeting a number of assumptions associated with the use of structural equation models.

1. Measurement error in any of the putative mediators can bias the results of a mediational analysis. If the "true" mediator is unreliably measured in a simple one mediator model (e.g., Fig. 2), the importance of that mediational path will be underestimated. Measurement error has much more complex effects when measurement error exists in multiple putative mediators in models such as that depicted in Fig. 3c. We are not guaranteed that the direction of bias will be downward (underestimating mediation). When measurement error exists in one of a set of correlated mediators, then the coefficients of all of the mediators may be biased (see discussions of measurement error in Aiken & West, 1991; Duncan, 1975; Kenny, 1979). This problem can be overcome by the use of structural equation models using multiple indicators of each of the measured constructs (Judd & Kenny, 1981b), though often at a cost of requiring a much larger sample size to achieve proper estimation of the more complex model.

2. A mediational analysis requires the specification of a small theory of the intervention. Mediational analysis presumes that small theory of the intervention is correct and examines the fit of the model implied by the small theory to the data. However, if the small theory is not correct (a condition known as a specification error), biased results may occur. Three potential problems are of special note. (a) The intervention may affect an important unmeasured third variable, which, in turn, causes both the mediator and the outcome. (b) The mediator(s) and the outcome may mutually influence each other (bidirectional causality). (c) The presence of one of the mediators may be a necessary condition for the operation of a second mediator. For example, the mediational analysis presented above based on data from Aiken et al. (1993) indicated that perceived benefits was the only variable that appeared to mediate the effect of the intervention on intentions to get a screening mammogram. However, these results may be misleading; a plausible alternative hypothesis is that women would be unlikely to be motivated to get an expensive screening mammogram unless they also perceived that they were susceptible to breast cancer. The perceived susceptibility component may be a necessary condition for perceived benefits to have its impact.

Each of these three potential problems can lead to serious mis-estimates of the role of the putative mediators in the model. Techniques for detecting some forms of specification errors in mediational analysis are discussed in James and Brett (1984). Design enhancements discussed below can also address some of these problems.

3. The test of condition 3 may include several putative mediators as well as the intervention in a regression equation predicting the outcome variable, as illustrated in Fig. 3c. Even when the model is correctly specified, multicollinearity among the set of mediators and the intervention will be created, particularly when the intervention has relatively strong effects on the mediators. The mediators and the intervention serve as predictors of the outcome variable in the regression equation; with high inter-predictor correlation there are difficulties in apportioning unique variance to the mediators and the intervention, large standard errors of the path coefficients, and low statistical power for the tests of mediation. Larger sample sizes can help alleviate this problem. Alternatively, the use of factorial or fractional factorial designs which include intervention components designed to target specific mediators can reduce the multicollinearity among the mediators.

4. Mediational analyses can be extended to investigate several forms of interactions that may occur. Tests of interactions of an intervention with the participant's level on a pretest measure (e.g., initial level of symptoms) can be accomplished with straightforward extensions of the techniques described above (see Baron & Kenny, 1986; James & Brett, 1984 for techniques; Wolchik et al., 1993 for an empirical example). More problematic are two other forms of interactions: (a) interactions between treatment components, and (b) interactions between mediators. These two forms of interactions often cannot be effectively probed when the prevalent treatment package strategy or the comparative treatment strategy designs are used in the randomized trial.

To illustrate, reconsider the hypothesis raised earlier that perceived susceptibility to breast cancer is a necessary condition for perceived benefits of mammography to have an impact on intentions to get a screening mammogram. This hypothesis predicts an interaction between perceived susceptibility and perceived benefits such that perceived benefits will be related to intentions only with perceived susceptibility is high. To the extent the intervention is effective, the level of both these putative mediators will be relatively high in the intervention condition and low in the control condition. This means that there will be few participants for whom perceived susceptibility is high, but perceived benefits are low or perceived susceptibility is low, but perceived benefits are high. The existence of such "off diagonal" cases is absolutely essential if stable estimates of interaction effects are to be produced. Strong tests of such interactions require that additional intervention conditions containing components that target only one of the two mediators be added to the design.

Despite these potential limitations, mediational analysis remains a promising, efficient method of probing the processes through which treatment components exert their effects on the outcome variable. These techniques can be applied to prevention trials involving more than two intervention conditions, multiple mediators, and extended causal chains. Theoretically, these techniques can also be extended to cases in which curvilinear relations among variables are expected. When the assumptions are met, these techniques provide strong tests of the small theory of the intervention and may have the potential to enhance the contribution of intervention trials to basic psychosocial research. Limited techniques exist for investigating the extent to which the assumptions are violated and in some cases for correcting for effects of these violations. In addition, several features may be added in the design of the intervention trial that can help minimize several of these potential problems.

CONCLUSION

At the beginning of this article we raised three questions about the results of trials of multicomponent interventions that are of concern to researchers. Question 1, identification of the influence of each component on the desired final outcome, is best addressed through the use of factorial designs, or fractional factorial designs if certain effects can plausibly be assumed to be negligible. The primary drawback in the application of these designs to interventions with multiple components is that large sample sizes may be required for adequate statistical power if higher order interactions between components are of interest. Question 2, identifying the combination of components that produces the optimal outcome tentatively appears to be best addressed by response surface methodology. Sample size requirements to achieve adequate statistical power may limit the applicability of this approach for multicomponent interactions. Further, the rarity of published examples raises issues about the success with which components can be scaled for treatment strength, a requirement of response surface methodology. Nonetheless, some adaptation of the basic approach of initially estimating the form of a response surface and then investigating that surface through sequential experimental trials would seem to hold considerable promise. Finally, question 3, the processes through which each component achieves its effect on the final outcome, is addressed through the use of mediational analyses. Such analyses require the specification of a small theory of the intervention which is rarely a feature of current reports of intervention trials. However, clear specification of the likely path(s) of influence of each component hold considerable promise for enhancing the basic science contribution of the results of intervention trials. Mediational analyses do have important potential limitations because of their correlational base, but many of these features can be addressed through focused analyses (e.g., statistical correction for error of measurement) or design enhancements (e.g., multiple indicators of each construct) to address specific problems.

A theme that clearly underlies nearly all of the topics discussed in this paper is the necessity to articulate clearly the effects that are of interest in both the design and analysis of the intervention trial. The use of only two levels of an intervention in the design strongly presumes that only a linear effect can occur; more than two levels are required to detect curvilinear effects. Similarly, the failure to include higher order (e.g., quadratic) components in regression or structural equation analyses means that only linear effects can be detected. Likewise, many of the techniques discussed in this article make strong assumptions about the nature of the effects of the components that are of interest. If these assumptions are seriously violated, the techniques will yield biased estimates of these effects. At the same time, readers should recognize that the traditional psychotherapy research designs such as the comparative and constructive designs also make strong assumptions that have rarely been articulated by researchers. Interventions representing combinations of components may be more or less than the sum of individually effective components. Researchers need to use the best technique to address their specific questions of interest and to state clearly the assumptions that have been made. However, designs and analyses should be developed to the extent possible to be capable of probing the plausibility of the assumptions that have been made, particularly if they are compatible with other competing theoretical viewpoints (Coie et al., 1993).

Because of space limitations in this article, we have not explored a number of hybrid design and analysis strategies that appear to be promising. One major class of these designs takes advantage of the temporal sequencing of the intervention components, the measurements, or both. For example, the components of multicomponent interventions are often introduced sequentially over the duration of a multi-week program. It may be possible to collect measures of each of the putative mediators and the outcome variable at the point of completion of each program component. Mediational analyses can then be used to probe the effect of each component on the outcome. Additional information about mediation may also be gleaned from analyses using measures of the mediators and outcome collected at the posttest and follow-up measurement waves. However, this strategy is often comprised by the popularity of wait list control designs in many preventive trials which contaminate the control group following the delayed intervention. Finally, designs in which a small subsample of individuals is *randomly* selected for intensive study from each of the conditions in the randomized trial can potentially yield strong information about the processes of change. Although the full range of design and analysis issues have not been outlined for this class of interventive designs, Stone, Kessler, and Haythornthwaite (1991) and West and Hepworth (1991) present discussions of many of these issues in intensive studies of daily experience.

In this article we have presented a number of design and analysis options, many of which have not been widely used in prevention research. Some of these will turn out to be of widely useful, having applications in prevention research beyond those envisioned here. Others may become useful only after some further adaptation or only in limited areas of application. The promise of this class of techniques that examine the role of individual components in prevention research is considerable. These techniques improve the construct validity of our interventions and may enhance the contribution of large scale prevention trials to our understanding of basic psychosocial development. These techniques also identify key program components and the processes necessary to produce favorable program outcomes, increasing our ability to successfully export good programs to new sites. Addressing such worthy basic science and applied science goals is likely to be a new and important focus of the next generation of prevention research.

ACKNOWLEDGMENT

The first author was partially supported by NIMH Grant P50MH39246 during the writing of this article. The research described in the section on mediational analysis was supported by National Cancer Institute Grant R03-CA46736. We thank Associate Editor Edward Seidman and four anonymous reviewers for their helpful suggestions for revision.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Aiken, L. S., & West, S. G. (April 1993). Outcome evaluation of community based programs to increase mammography screening. In L. S. Aiken and S. G. West (Chairs), Development, Implementation, and Evaluation of Interventions to Increase Mammography Screening: A Symposium. Western Psychological Association, Phoenix, AZ.

Aiken, L. S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology, and measurement in psychology: A survey of PhD programs in North America. *American Psychologist, 43,* 721-734.

Aiken, L. S., West, S. G., Woodward, C. K., Reno, R. R., & Reynolds, K. R. (1993). *Increasing screening mammography in asymptomatic women: Evaluation of a second generation, theory-based program.* Manuscript submitted for publication.

Anderson, V. L., & McLean, R. A. (1984). *Applied factorial and fractional designs.* New York: Marcel Dekker.

Atkinson, A. C. (1985). An introduction to the optimum design of experiments. In A. C. Atkinson and S. E. Fienberg (Eds.), *A celebration of statistics: The ISI centenary volume.* New York: Springer-Verlag, pp. 465-473.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-606.

Borenstein, M., & Cohen, J. (1988). *Statistical power analysis: A computer program.* Hillsdale, NJ: Erlbaum.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces.* New York: Wiley.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building.* New York: Wiley.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Caplan, R. D., Vinokur, A. D., Price, R. H., & Van Ryn, M. (1989). Job seeking, reemployment and mental health. *Journal of Applied Psychology, 74,* 759-769.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coie, J. D., Watt, N., West, S. G., Hawkins, D., Asarnow, J., Markman, H., Ramey, S., Shure, M., & Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist, 48,* 1013-1022.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton-Mifflin.

Cooper, W. H., & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology, 71,* 179-184.

Duncan, O. D. (1975). *Introduction to structural equation models.* New York: Academic.

Durlak, J., Wells, A., Cotten, J., & Lampmann, C. (June 1993). A review of primary prevention programs for children and adolescents. In J. Durlak (Chair), Evaluating Primary Prevention: Programs, Outcomes, and Issues. Symposium presented at the Fourth Biennial Conference on Community Research and Action, Williamsburg, VA.

Fienberg, S. E., Singer, B., & Tanur, J. M. (1985). Large-scale social experimentation in the United States. In A. C. Atkinson and S. E. Fienberg (Eds.), *A celebration of statistics: The ISI centenary volume*, New York: Springer-Verlag, pp. 287-326.

Flay, B. R. (1985). Psychosocial approaches to smoking prevention: A review of findings. *Health Psychology, 4,* 449-488.

Greenwald, A. G. (1976). Within-subject designs: To use or not to use? *Psychological Bulletin, 83,* 314-320.

Hansen, W. B., Johnson, C. A., Flay, B. R., Graham, J. W., & Sobel, J. (1988). Affective and social influences approaches to the prevention of multiple substance abuse among seventh grade students: Results from Project SMART. *Preventive Medicine, 17,* 135-154.

Harackiewicz, J. M., Sansone, C., Blair, L. W., Epstein, J. A., & Manderlink, G. (1987). Attributional processes in behavior change and maintenance: Smoking cessation and continued abstinence. *Journal of Consulting and Clinical Psychology, 55,* 372-378.

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin, 112,* 64-105.

Higginbotham, H. N., West, S. G., & Forsyth, D. R. (1988). *Psychotherapy and behavior change: Social, cultural, and methodological perspectives.* New York: Pergamon.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69,* 307-321.

Judd, C. M., & Kenny, D. A. (1981a). *Estimating the effects of social interventions.* New York: Cambridge University Press.

Judd, C. M., & Kenny, D. A. (1981b). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5,* 602-619.

Kazdin, A. E. (1980). *Research design in clinical psychology.* New York: Harper & Row.

Kazdin, A. E. (1986). The evaluation of psychotherapy: Research design and methodology. In S. L. Garfield and A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed.), New York: Wiley, pp. 23-68.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

Kreft, I. (August 1992). Hierarchical Linear Models: Potential Applications in Psychological Research. Invited Address at the meeting of the American Psychological Association, Washington, D.C.

Lipsey, M. W. (1992). Theory as method: Small theories of treatments. In L. B. Sechrest and A. G. Scott (Eds.), *New directions for program evaluation* (No. 57). San Francisco: Jossey-Bass, pp. 5-38.

Maccoby, N., & Farquhar, J. W. (1975). Communication for health: Unselling heart disease. *Journal of Communication, 25,* 114-126.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediating effects in prevention studies. *Evaluation Review, 17,* 144-158.

MacKinnon, D. P., Johnson, C. A., Pentz, M. A., Dwyer, J. H., Hansen, W. B., Flay, B. R., & Wang, E. (1991). Mediating mechanisms in a school-based drug prevention program: First year effects of the Midwestern Prevention Project. *Health Psychology, 10,* 164-172.

Mason, R. L., Gunst, R. F., & Hess, J. L (1987). *Statistical design and analysis of experiments: With applications to engineering and science.* New York: Wiley.

Mead, R. (1988). *The design of experiments: Statistical principles for practical application.* Cambridge: Cambridge University Press.

Myers, R. H., Khuri, A. I., & Carter, W. H., Jr. (1989). Response surface methodology: 1966-1988. *Technometrics, 31,* 137-157.

Pentz, M. A., Dwyer, J. H. MacKinnon, D. P., Flay, B. R., Hansen, W. B., Wang, E. Y. I., & Johnson, C. A. (1989). A multi-community trial for primary prevention of adolescent drug abuse: Effects on drug use prevalence. *Journal of the American Medical Association, 261,* 3259-3266.

Perri, M. G., McAllister, D. A., Gange, J. J., Jordan, R. C., McAdoo, W. G., & Nezu, A. M. (1988). Effects of four maintenance programs on the long-term management of obesity. *Journal of Consulting and Clinical Psychology, 56,* 529-534.

Pilz, J. (1983). *Bayesian estimation and experimental design in linear regression models.* Leibzig: Teubner-Texte.

Reynolds, K. D., West, S. G., & Aiken, L. S. (1990). Increasing the use of mammography: A pilot program. *Health Education Quarterly, 17,* 429-441.

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting and Clinical Psychology, 58,* 646-656.

Sandler, I. N., West, S. G., Baca, L., Pillow, D. R., Gersten, J. C., Rogosch, F., Virdin, L., Beals, J., Reynolds, K. D., Kallgren, C., Tein, J.-Y., Kriege, G., Cole, E., & Ramirez, R. (1992). Linking empirically-based theory and evaluation: The Family Bereavement Program. *American Journal of Community Psychology, 20,* 491-521.

Sechrest, L., West, S. G., Phillips, M. A., Redner, R., & Yeaton, W. (1979). Some neglected problems in evaluation research: Strength and integrity of treatments. In L. Sechrest and associates (Eds.), *Evaluation studies review annual* (Vol. 4), Beverly Hills, CA: Sage, pp. 15-35.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect of the power of studies? *Psychological Bulletin, 105,* 309-316.

Shadish, W. R. (1992). The logic and design of randomized experiments. Department of Psychology, Memphis State University, Unpublished manuscript; (chapter to appear). In T. D. Cook, W. R. Shadish, and D. T. Campbell (Eds.), *Quasi-experimentation: Design and analysis issues for field settings* (2nd Ed.). Boston: Houghton-Mifflin.

Shaffer, D., Philips, I., & Enzer, N. B. (Eds.) (1989). Prevention of Mental Disorders, Alcohol and Other Drug Use in Children and Adolescents. U.S. Department of Health and Human Services, Office for Substance Abuse Prevention, Rockville, MD, DHHS Publication No. (ADM) 90-1646.

Shure, M. B. (1988). How to think, not what to think: A cognitive approach to prevention. In L. A. Bond and B. M. Wagner (Eds.), *Families in transition: Primary prevention programs that work.* Newbury Park, CA: Sage, pp. 170-199.

Silvey, S. D. (1980). *Optimal design.* New York: Chapman & Hall.

Steinberg, D. M., & Hunter, W. G. (1984). Experimental design: Review and comment (with discussion). *Technometrics, 26,* 71-130.

Stone, A. A., Kessler, R. C., & Haythornthwaite, J. A. (1991). Measuring daily events and experiences: Decisions for researchers. *Journal of Personality, 59,* 575-608.

Tharp, R. G., & Gallimore, R. (1979). The ecology of program research and evaluation: A model of evaluation succession. In L. Sechrest and Associates (Eds.), Evaluation studies review annual (Vol. 4). Beverly Hills, CA: Sage, pp. 39-60.

Webster-Stratton, C., Kolpacoff, M., & Hollinsworth, T. (1988). Self-administered videotape therapy for families with conduct-problem children. *Journal of Consulting and Clinical Psychology, 56,* 558-566.

Weissberg, R. P., Caplan, M. Z., & Sivo, P. J. (1989). A new conceptual framework for establishing school-based social competence promotion programs. In L. A. Bond and B. E. Compas (Eds.), *Primary prevention and promotion in the schools,* Newbury Park, CA: Sage, pp. 255-296.

West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality, 59,* 609-662.

West, S. G., Newsom, J. T., & Fenaughty, A. M. (1992). Publication trends in JPSP: Stability and change in the topics, methods, and theories across two decades. *Personality and Social Psychology Bulletin, 18,* 473-484.

Whalen, C. K., Henker, B. Swanson, J. M., Granger, D., Kliewer, W., & Spencer, J. (1987). Natural social behaviors in hyperactive children: Dose effects of methylphenidate. *Journal of Consulting and Clinical Psychology, 55,* 187-193.

Wolchik, S. A., West, S. G., Westover, S., Sandler, I. N., Martin, A., Lustig, J., Tein, J.-Y., & Fisher, J. (1993). The children of divorce intervention project: Outcome evaluation of an empirically based parenting program. *American Journal of Community Psychology, 21,* 293-331.

Woodward, J. A., Bonett, D. G., & Brecht, M.-L. (1990). *Introduction to linear models and experimental design.* San Diego, CA: Harcourt Brace Jovanovich.