



For Violence Prevention

Revised 11/17/11

## BLUEPRINTS FOR VIOLENCE PREVENTION SELECTION CRITERIA

The success of a community's violence prevention efforts will depend, in large degree, upon the preventive interventions used. That is why it is imperative to identify approaches that have been proven effective. Although a program model can rarely, if ever, be proven to be superior to all others, a particular model elicits greater confidence after its theoretical rationale, goals and objectives, and outcome evaluation data have been carefully reviewed. Although various scholarly reviews have identified exemplary programs, the methodological standards used in evaluating program effectiveness can vary. A few of these scholarly reviews have explicit standards, and a few score each program evaluation on its methodological rigor, but for most the standards are variable, sometimes unrelated to effectiveness, and seldom made explicit. The standard for the claims of program effectiveness in most of these reviews is very low. Of those with explicit standards, Blueprints programs have the highest standards and meet rigorous tests of effectiveness in the field. There are several important criteria to consider when reviewing program effectiveness. Three of these criteria are given greater weight: evidence of deterrent effect with a strong research design, sustained effect, and multiple site replication. Blueprints model programs must meet all three of these criteria, while promising programs must meet at least the first criterion.

### **Evidence of Effect with a Strong Research Design**

This is the most important of the selection criteria. Relatively few programs have demonstrated effectiveness in reducing the onset, prevalence, or individual offending rates of *violent behavior*. The Blueprints Advisory Board has historically accepted evidence of effects for three key indicators—violence (including childhood aggression and conduct disorder), delinquency, and/or drug use—as evidence of program effectiveness. Beginning in 2011, the Board expanded the outcomes to include mental health (anxiety, depression, suicide, self-regulation), educational skills and attainment, and some physical health outcomes.

Providing sufficient quantitative data to document effectiveness in preventing or reducing the above behaviors requires the use of evaluative designs that provide reasonable confidence in the findings (e.g., experimental designs with random assignment or quasi-experimental designs with matched control groups). Most researchers recognize random assignment studies (randomized trials) executed with fidelity as providing the highest standard of program evaluation. Random assignments offer the most compelling evidence that study results are due to the intervention rather than to preexisting differences between experimental and control groups and/or other threats to internal validity, such as maturation, selection bias, and testing effects. In these studies, assignment to experimental or control conditions is determined solely by chance, and the likelihood of differences being attributed to the assignment process can be assessed.

When random assignment cannot be used, the Advisory Board considers studies that use control groups matched as closely as possible to experimental groups on relevant characteristics (e.g.,

gender, race, age, socioeconomic status, income) and studies with control groups that use statistical techniques to control for initial differences on key variables. As carefully as experimental and control groups are matched, however, it is impossible to determine if the groups may vary on some characteristics that have not been matched or controlled for and that are related to program outcome. Random assignment, therefore, is believed to be the most rigorous of methodological approaches.

Research designs vary greatly in quality, particularly with respect to several key aspects: sample size and selection, attrition (loss of study participants over time), measurement, and analysis issues. At a minimum, the following issues need to be addressed:

(1) Sample sizes must be large enough to provide statistical power to detect at least moderate sized effects. It is more difficult to detect statistically significant differences between groups when small sample sizes are used. Selection of participants must be made in a manner that avoids bias. For example, a self-selecting sample that relies on volunteer participants might be more motivated to make change, thus introducing a plausible alternative explanation for outcomes that are achieved. An adequate description should report the characteristics of the sample, the selection process, and pretest differences on relevant variables between the treatment and control conditions.

(2) Attrition, or loss of study participants, may indicate problems in program implementation or failure to locate subjects during a follow-up period. Attrition is dangerous, particularly because it can compromise the integrity of the original randomization or matching process. It reduces confidence that the original sample and final sample are comparable and that the final experimental and control comparisons reflect only treatment effects. Sample sizes and losses must be reported through all follow-up periods, and tests that rule out differential attrition should be conducted.

(3) Tests to measure outcomes must be administered fairly, accurately and consistently to all study participants. For example, the use of inconsistent measures over time may produce less reliable test scores. The instruments used to measure outcomes should be demonstrated to be reliable and valid. Measurements of actual behavior are required for Blueprints, not attitudes or intent. More than one report of behavior is preferable in instances where the same person both delivers the intervention and provides a measure of the outcome. For example, a family intervention that teaches mom how to interact with her child should not rely solely on mom's report of the child's behavior. When multiple measures of outcomes are used in a study, the intervention should significantly influence the most important outcomes and influence the others in the expected direction.

(4) Analyses should be appropriately designed. They should be done at the same level as the randomization and, following an "intent to treat" approach, should include all participants originally assigned to treatment and control conditions. Secondary analyses can be performed to determine the effectiveness of a program at differing levels of implementation and dosage. Two-tailed tests of significance are preferred since they represent the most conservative of tests.

***School-based Evaluations.*** Evaluations of school-based programs, with schools as the unit of analysis, typically require multiple schools per condition to perform a main effects analysis with sufficient power to detect effects. Since meeting this criterion requires a complex and costly evaluation, it would eliminate most existing school-level studies from consideration in the Blueprints Series. Therefore, school-based evaluations that use experimental or quasi-experimental designs with relatively few schools, but *more* than one in each condition, will be considered in the Blueprints Series if they meet an additional burden of proof. They must demonstrate consistency across effects and across replications with multiple measures from different sources. The theoretical rationale should be well developed, and there should be a rigorous evaluation of theory with evidence that the results are consistently in line with the expectations (i.e., there are changes in the risk and protective factors which mediate the changes in outcomes). Outcomes should be robust, with at least moderate effect sizes. Evidence that the benefits of the program outweigh the costs is helpful. The decision to accept this level of proof is driven entirely by the state of current research. It should not be assumed that this standard of proof is ideal. Evaluations with multiple schools are most desirable and should be encouraged among funders and researchers.

### **Sustained Effects**

Although one criterion of program effectiveness is that it demonstrates success by the end of the treatment phase, it is also important to demonstrate that these program effects endure beyond treatment and from one developmental period to the next. Designation as a Blueprints model program requires a sustained effect at least one year beyond treatment, with no subsequent evidence that this effect is lost. Unfortunately, many programs that demonstrate initial success fail to show long-term maintenance of the effects after the intervention has ended. Depending on whether effects are immediate or delayed, the full impact of an intervention or treatment may not be realized by the end of treatment. Significant improvement may be realized over time, or a decay or decline may result. For example, if a preschool program designed to offset the effects of poverty on school performance (e.g., Head Start) demonstrates its effectiveness when children start school, it is also important to demonstrate that these effects are sustained over a longer period of time. Unless this protective effect is sustained through high school, it is unlikely to have an impact during this critical period when problem behavior is at its peak: a sustained effect will most help adolescents maintain a successful life course trajectory.

A program may be identified as promising without meeting the sustainability criterion. In some cases, programs may not have conducted longer-term follow-ups. In other cases, programs will have performed long-term follow-ups and found no enduring effects. If program effects disappear at a later time period, Blueprints may qualify the program for only the period of time in which it was found to be effective, stating the loss of enduring effects at the point at which they were found. While these programs may not show enduring effects for 12 months or longer on specifically measured outcomes, in some cases they can provide meaningful benefits to youth, schools, and communities. For example, even if benefits don't last, delaying the onset of alcohol and drug use to a later age would improve the safety of youth during a highly vulnerable period of their lives. And since early onset of youth problems often leads to more serious problems later, delaying onset with temporary improvements may have payoffs at older ages.

## Replication

Replication is an important element in establishing program effectiveness and understanding what works best, in what situations, and with whom. Some programs are successful because of unique characteristics in the original site that may be difficult to duplicate in another site (e.g., having a charismatic leader or extensive community support and involvement). Replication establishes the strength of a program and its prevention effects and demonstrates that it can be successfully implemented in other sites. Blueprints considers replication to be synonymous with dependability.

Programs that have demonstrated success in diverse settings (e.g., urban, suburban, and rural areas) and with diverse populations (e.g., different socioeconomic, racial, and cultural groups) create greater confidence that such programs can be transferred to new settings. As communities prepare to tackle problems of violence, delinquency, and substance abuse, knowledge that a specific program has had success in varied settings with similar populations adds to its credibility.

Some projects may be initially implemented as a multisite single design (i.e., several sites are included in the evaluation design). Although not as valuable as independent replications, these designs can check for overall main effects and sources of variation across sites. Becoming a Blueprints model program requires at least one high-quality replication with fidelity demonstrating that the program continues to be effective. This criterion does not need to be met to qualify as a promising program.

Replication dismantling designs will also be considered. If a program has been implemented and evaluated as a component within a number of different programs (multiple component studies) and has also been implemented and evaluated alone, it is possible that the multiple component studies might meet the replication criterion. There must be a total of three studies, including the standalone program evaluation and two additional multiple component studies. All must be well designed with positive effects and with no negative effects. A hypothetical example of this type of replication would be as follows:

Study 1: Good Behavior Game alone

Study 2: Good Behavior Game + Classroom Centered Intervention

Study 3: Good Behavior Game + Family Classroom Partnering

## Additional Factors

In the selection of Blueprints model programs, two additional factors are considered: whether a program conducted an analysis of mediating factors and whether a program is cost effective.

***Analysis of Mediating Factors.*** The Blueprints Advisory Board looks for evidence that change in the targeted risk or protective factor(s) mediates the change in problem behaviors. This evidence clearly strengthens the claim that participation in the program is responsible for the change in behavior, and it contributes to our theoretical understanding of the causal processes involved. In its reviews of different programs, the Advisory Board has discovered that many

programs reporting significant outcome benefits have not collected the data necessary to complete an analysis of mediating factors.

**Costs versus Benefits.** Program costs should be reasonable and should be less or no greater than the program's expected benefits. High price-tag programs are difficult to sustain when competition is high and funding resources low. Implementing expensive programs that will, at best, have small effects on violence is counter-productive. Although outcome evaluation research established that Blueprints programs were effective in reducing violence, delinquency, and drug use, very few data were available initially regarding the costs associated with replicating these programs.

Two cost-benefit studies involving Blueprints programs—the RAND Corporation Study and a study by the Washington State Institute for Public Policy--suggest that these programs are cost-effective (Greenwood, Model, Rydell, & Chiesa, 1996; Washington State Institute for Public Policy, 2011).

**Summary.** The selection criteria identified above establish a high standard, one that has proved difficult for most programs to meet, thus explaining the small number of Blueprints programs. This high standard reflects the level of confidence necessary, however, for recommending that communities replicate these programs with reasonable assurances that they will prevent violence and other behavioral problems when implemented with fidelity. The Blueprints model programs are not intended to be a comprehensive list of programs that work, but rather reflect a selection of programs with strong research designs for which we have found good evidence of their effectiveness. There is no implication that programs *not* on this list are necessarily ineffective. Chances are that there are a number of good programs that have just not yet undergone the rigorous evaluations required to demonstrate effectiveness. But our evaluations have also revealed that many programs are ineffective, and a few are iatrogenic (i.e., harmful). Without evaluations, we just don't know. It is in the best interests of our children to evaluate, so we can have confidence that what we are doing for them actually helps. As time goes on and new research findings are published, Blueprints will add to this list other credible, effective programs which communities can use confidently. It will also continue to follow evaluations of Blueprints programs to refine our knowledge of their effectiveness for specific populations and over longer periods of time.