THE IMPORTANCE OF IMPLEMENTATION FIDELITY

by Sharon Mihalic
Director, Blueprints for Violence Prevention Initiative
Center for the Study and Prevention of Violence
442 UCB
Boulder, Colorado
April, 2002

Over the past several years, a large amount of information has been collected on the risk and protective factors for violence. Research has also identified prevention programs that can modify these risk and protective factors. The Blueprints initiative has been in the forefront in identifying exemplary programs that have been evaluated in rigorous, controlled trials, and much attention has been focused nationally on selecting and implementing quality programs. However, identification of effective programs is only the first step in the efforts to prevent and control violence. Widespread implementation of effective programs is unlikely to affect the incidence of violent crime unless there is careful attention given to the quality of implementation, the degree to which a program is delivered as intended (American Youth Policy Forum, 1999; Biglan & Taylor, 2000; Lipsey, 1999). Research demonstrates that successful implementation is not guaranteed by a site's decision to adopt a best practices program. Many science-based programs have been adopted in different settings with widely varying outcomes. In fact, a high quality implementation of a poor program may be more effective than a low quality implementation of a best practice program (Gottfredson, Gottfredson, & Czeh, 2000). Until recently, little emphasis has been given to implementing programs with fidelity in both the science and practice of prevention. As a result, most people do not recognize the importance of implementation fidelity and feel that implementation of at least some program components will be better than doing nothing. However, this may be an erroneous belief, since we typically do not know which components of a program may be responsible for the reductions in violence. Programs must be implemented with fidelity to the original model to preserve the behavior change mechanisms that made the original model effective (Arthur & Blitz, 2000).

Defining Implementation Fidelity

The concept of implementation fidelity, sometimes called adherence or integrity, is a determination of how well the program is being implemented in comparison with the original program design (i.e., is the program being delivered as it was designed and implemented in its efficacy and/or effectiveness trials). There are four primary components examined when considering program fidelity (Dane & Schneider, 1998):

1. *Adherence* refers to whether the program service or intervention is being delivered as it was designed or written, i.e., with all core components being delivered to the appropriate population; staff trained appropriately; using the right protocols, techniques, and materials; and in the locations or contexts prescribed;

2. *Exposure* may include any of the following: the number of sessions implemented, length of each session, or the frequency with which program techniques were implemented;

3. *Quality of Program Delivery* is the manner in which a teacher, volunteer, or staff member delivers a program (e.g., skill in using the techniques or methods prescribed by the program, enthusiasm, preparedness, attitude); and

4. *Participant Responsiveness* is the extent to which participants are engaged by and involved in the activities and content of the program.

Although the concept of implementation fidelity is not new, ways in which to operationalize, or measure fidelity is a relatively recent phenomenon. Fidelity is assessed by conducting a process evaluation. Appendix A contains a detailed definition and describes why a process evaluation should be conducted when implementing a program. It also contains the major elements of the process evaluation that was conducted for the Blueprints replication sites.

Importance of Implementation Fidelity

Although an extremely important topic, program implementation has been relatively neglected in the prevention research literature (Fagan, 1990; Greenberg et al., 2001). In a review of over 1,200 published prevention studies, only 5 percent provided data on implementation (Durlak 1997) . In a review of 34 rigorously evaluated programs to prevent mental disorders in school-age children, 11 of the 34 studies (32 percent) utilized implementation information as a source of data for outcome analyses, and a third of these linked variability in implementation indices to differences in program outcomes (Domitrovich & Greenberg, 2000). Dane and Schneider (1998) found that only 39 of the 162 preventive interventions they examined contained information on program integrity, and only 13 of these studies considered the impact of fidelity on outcomes. Another examination of 181 experimental studies published between 1980-90 in seven journals known for behaviorally based interventions was reviewed, and only 14.4% of the studies had systematically measured and reported integrity data; only 34% had operationally defined treatments (Gresham et al., 1993).

When evaluations do examine program fidelity, many studies have found that they are not being implemented with strength and fidelity to the original model, although several hallmark studies of health programs have underscored the importance of the quantity and quality of implementation (Connell, Turner, & Mason, 1985; Taggart, Bush, Zuckerman, & Theiss, 1990; Resnicow, Cohn, Reinhardt, Cross, Futterman, Kirschner, Wynder, & Allegrante, 1992). An example comes from the National Study of Delinquency Prevention in Schools in which several criteria were applied to discretionary prevention activities. The following represents the average level of intensity and fidelity to good prevention practice (Gottfredson, Gottfredson, & Czeh, 2000):

. One or more persons is conducting the prevention activity(ies) *from time to time*;
. It employs 71% of the *content* elements identified as representing best practices;
. It employs 54% of the *methods* elements identified as representing best practices;
. It involves 32 sessions or lessons (although there is a large range across activities of different types);

. It lasts about 25 weeks;

. Both students and staff participate about once per week;

. 41% of the school's students participate or are exposed;

. There are approximately 4 program providers per 100 students in the school; and

. If it is a classroom or a school-wide activity, it operates nearly all year.

The researchers found that the average prevention activity (i.e., strategy) received a passing grade on only 57% of the quality criteria examined (i.e., only 57% of the indicators of quality or quantity were judged to be sufficiently strong enough to lead to behavior change), thus concluding that the quality of prevention activities in the nation's schools is generally poor and that prevention activities are not being implemented with sufficient strength and fidelity to be expected to produce a measurable difference in the desired outcomes.

In another example, a study that examined school-based programs sponsored by the Department of Education's Safe and Drug Free Schools and Communities Program found that these programs were not implemented with the same attention to core components and dosage as found in the research models (Silvia & Thorne, 1997).

Ensuring that community providers understand the core program components and dosage that are necessary for success is a serious challenge to program developers and disseminators. The original trials (i.e., efficacy studies) of programs are usually under the maximum control of the designer and under optimal conditions with high levels of funding, motivation, and support. The researcher generally exercises extreme care to ensure that the program is thoroughly understood and implemented with a high degree of quality. As programs are proven effective and disseminated widely, in naturalistic settings under less favorable conditions (effectiveness studies), the chances for key program components to be modified and inconsistencies in program delivery become more likely (Dane & Schneider, 1998). Depending upon the type of modifications that are made, the program may become less effective in preventing the outcomes sought. In fact, the less involved

the researcher in the design, planning, and delivery of the intervention, the smaller the effect size on behavioral outcomes (Lipsey, 1999).

Modifications may be made by some practitioners with full knowledge of the program in an effort to adapt the program to fit local needs. In other instances, adaptations may be made because the site does not have a thorough understanding of the program and its underlying causal mechanism. Gresham et al. (1993) found that only 35% of the studies they reviewed even provided an operational definition of the intervention through a detailed description or reference to a manual. However, even when detailed descriptions and manuals are provided, implementation may still fall short of the ideal. For instance, a study to test the effectiveness of the Life Skills Training program in 56 New York State schools (Botvin, Baker, Dusenbury, Tortu, & Botvin, 1990) showed that only 27% to 97% (mean of 68%) of the material in the curriculum was covered, with only 75% of the students in the prevention conditions exposed to 60% or more of the prevention program (i.e., one in four students had teachers who implemented less than 60% of the important points of the lessons).

Because efficacy studies generally include youth that are receiving all program components, we can only conclude that a program works if implemented in its entirety. If specific components are omitted, it is possible that the omitted component could be the mechanism that is affecting much of the change in behavior, resulting in a loss of program effects. For example, the research design of the Adolescent Alcohol Prevention Trial (Donaldson et al., 1994) divided students into four groups. The students received either *information about consequences of drug use* only, *resistance skills* only, *normative education* only, or *resistance skills training* in combination with *normative education*. Refusal skills alone were not predictive of later substance use, although normative education alone was. Tests of the combined effects showed that the combination of resistance skills training and normative education resulted in the lowest rates of drug use. A school that fails to implement the lessons around both components, or implements refusal skills alone, might be making little or no impact in drug prevention or reduction, basically wasting time and money.

Meta-analyses also demonstrate that better implemented programs produce more change (Gresham et al., 1993; Wilson & Lipsey, 2000). Lipsey (1999) found that the best interventions can reduce recidivism by about 40 percent. Thorough implementation, however, was found to be a significant factor in relation to recidivism effects. Intervention effects were larger when attention was given to the integrity of the program implementation. Additionally, programs of more than six months' duration were, on average, more effective than those of shorter length. Table 1 shows the independent contribution of several program characteristics, described below, to recidivism rates. The comparison is the recidivism rate of routine probation, or treatment-as-usual services found in the control groups of these 200 studies. The base rate of 50 percent approximates that found in these control groups. The figure shows the successive decreases in recidivism if a minimal program (programs found to have smaller effect sizes in the meta-analysis, incomplete implementation, and less than six months duration) is added to routine services, then if that minimal program is upgraded to a more effective intervention (with larger effect sizes), if the program is thoroughly implemented, and with a longer duration. Similar results were also found for institutionalized offenders.

Table 1. Expected Recidivism with Various Intervention Characteristics for Noninstitutionalized Offenders

Intervention Characteristics	Recidivism
Routine Probation (P)	50%
P + Minimal Program	46%
P + Best Intervention Type (B)	40%
P + B + Good Implementation (I)	35%
P + B + I + Over 6 Months' Duration	32%

Evaluations of prevention programs can lead to conclusions that specific programs do not work when, in fact, the failure to find treatment effects may be the direct result of weaknesses in

program implementation. Moderate positive correlations have been found between degree of treatment integrity and level of treatment outcome (Gresham et al., 1993). Additionally, treatment integrity has been demonstrated to be important in changing the mediating variables that are theorized to be responsible for the outcomes (Hansen, Graham, Wolkenstein, & Rohrbach, 1991). Most outcome studies that have examined implementation have used measures of adherence and/or exposure. Fewer studies have examined the quality of program delivery and participant responsiveness.

Studies Examining Adherence

When a fidelity of implementation analysis (i.e., process evaluation) is conducted, it almost consistently shows superior outcomes when the program has been implemented with high fidelity (Fors and Doster, 1985; Gray, Emshoff, Jakes, and Blakely, draft; Gresham et al., 1993; McGrew, Bond, Dietzen, and Salyers, 1994). For example, analyses typically yield stronger prevention effects for the high-fidelity samples than for the full samples (Blakely et al., 1987; Dane & Schneider, 1998; Gottfredson, Gottfredson, & Hybl, 1993). In an evaluation of the Life Skills Training program (Botvin, Baker, Dusenbury, Botvin, & Diaz, 1995), results in the full sample indicated that the prevalence of heavy drinking, and weekly and monthly cigarette smoking was significantly lower for the intervention groups than the control group, and heavy smoking was significantly lower in one of the intervention groups than the control group. There were, however, no significant differences for the monthly, weekly, or 3 drinks or more per occasion rates, nor were there significant differences for marijuana use. In contrast, in the high fidelity sample, the results were stronger, and more outcomes became significant. The experimental groups were significantly lower than the control group for all measures of cigarette use, weekly alcohol use, 3 drinks or more per occasion, drunk, weekly marijuana use, monthly marijuana use, and monthly alcohol use.

In the Violent Juvenile Offender Program to test correctional interventions for chronically violent juvenile offenders, the two sites with stronger implementation of the program design (i.e., most of

the core components of the program received moderate to strong implementation ratings) fared much better than two sites that had not implemented the program well (these two programs received weak to moderate ratings for most of the core components). The well-implemented programs resulted in significant reductions in the number and severity of arrests for experimental youths compared to control youths, as well as in significantly greater time until rearrest (Fagan, 1990).

An evaluation of 13 Massachusetts Intensive Supervision Programs compared the degree of implementation with variation in outcomes across sites. This evaluation found that the more fully the ISP program was implemented, the more likely recidivism decreased significantly across a range of alternative outcome measures (Byrne and Kelly, 1989).

An intensive case study of the Positive Action Program was conducted in a rural school in northern Florida in the first year of implementation (Flay, 2000). The program was fully implemented in 11 classrooms, partially implemented in 7 classrooms, and sporadically or not implemented in 7 classrooms. Data was obtained from teachers, students, and parents at the beginning and end of the school year. Overall, teachers who had implemented more of the program improved more in their attitudes about and perceptions of other teachers, their own teaching effectiveness, and parent involvement. Students who received more of the program improved their positive attitudes and behaviors and decreased their negative attitudes and behaviors, such as disciplinary referrals, substance use, and violence. An increased level of implementation also improved parents readiness to take responsibility for their child's character and behavioral development, decreased their rating of their child's likelihood of giving into peer pressure, increased the level of communication with their child, and improved their knowledge of their child's friends and their parents. Furthermore, these effects were significant for students from both high and low socioeconomic status. Interestingly, the data suggests that receiving some or even most of PA is not sufficient for low income students, and that these students need the complete program for substantial effects to occur.

Multisystemic Therapy, a home-based family therapy program that targets violent or chronic juvenile offenders, has demonstrated in randomized trials remarkable reductions in rearrests, incarceration, self-reported offending, and a variety of effects on mediating variables related to family and peer relations, family functioning, and parental monitoring. However, in one study of violent and chronic juvenile offenders and their families, which omitted ongoing treatment fidelity checks, adolescent symptomatology and days incarcerated were reduced, but there was a lack of significant effects on criminal behavior and other instrumental (mediating) outcomes. In cases where treatment adherence ratings were high, the outcomes were substantially better (Henggeler, Melton, Brondino, Scherer, & Hanley, 1997). Another sample of substance-abusing juvenile offenders and their primary caregivers showed similar results (Huey, Henggeler, Brondino, & Pickrel, 2000). Furthermore, MST adherence had both a direct impact on delinquent behavior, as well as an influence that was partially mediated through its effects on family functioning and cohesion, parent monitoring, and delinquent peer association (Huey et al., 2000).

Some programs only have significant effects in the high-fidelity samples. For example, the Child Development Program was evaluated in 12 program schools, however, only five of these schools showed clear evidence of widespread program implementation. There was no clear evidence of positive program outcomes for students at all 12 program schools; however, at the five high-fidelity schools, there were significant declines in alcohol and marijuana use and an increase in students' sense of school as a community, compared to control schools (Battistich, Schaps, Watson, Solomon, & Lewis, in press).

A study of the Life Skills Training program in eight urban New York schools showed that the effects of the program were due only to the high implementation teachers with a mean completion rate of 78% of the material. The low implementation teachers implemented the material with a mean of 56% (Botvin, Dusenbury, Baker, James-Ortiz, & Kerner, 1989). Another study of the Life Skills Training Program in ten suburban New York junior high schools showed no significant differences between the teacher-led implementation and the control group in the full sample on smoking, alcohol, and marijuana use (Botvin, Baker, Filazzola, and Botvin, 1990). (It should be

noted that a peer-led booster condition did reduce substance use.) Even worse, the control group had significantly fewer drinkers than the teacher-led group on several of the alcohol measures. However, in a restricted sample of teachers who had implemented the program with a reasonable degree of fidelity (i.e., teachers who received a rating of 4 or 5 on a scale ranging from 1 to 5 in terms of implementation fidelity), the proportion of smoking, alcohol, and marijuana users declined. In one other Life Skills Training study (Botvin et al., 1990), there were no positive effects found in the sample of teachers who exposed students to less than 60% of the material. By using a three-year cumulative implementation score of 60% as the inclusion criteria for the analysis of program effectiveness, significant treatment effects were found for three of five measures of substance use.

A test of a theory-based intervention (Seattle Social Development Project) that seeks to change the opportunity, skill and reinforcement structures of mainstream classrooms by training teachers in educational strategies designed to promote academic achievement and school bonding showed that it was only through thorough implementation of these teacher practices that students' levels of classroom opportunity, involvement, reinforcement, and bonding to school was increased (Abbott, O'Donnell, Hawkins, Hill, Kosterman, & Catalano, 1998).

Modifying or adding components to a program can also present a serious threat to program fidelity, especially if the modification or addition consists of elements that have been found to be harmful to youth. For example, the use of scare tactics in drug or violence prevention have not been found to be effective (Botvin, 1990; Hansen, Graham, Wolkenstein, Lundy, Pearson, Flay, & Johnson, 1988) and in some cases have been found to be harmful (Petrosino, Turpin-Petrosino, & Finckenauer, 2000). Efforts to introduce these elements into already proven programs may backfire and result in a reduction of the program benefits that might have otherwise been expected. In the Midwestern Prevention Project (Pentz, Trebow, Hansen, MacKinnon, Dwyer, Johnson, Flay, Daniels, & Cormack, 1990), while none of the teachers reported that they had deviated from the program substantially, 68% of the teachers deviated slightly. All deviations were to include additional material, discussion, or sessions to the program. These deviations had

no significant effects on cigarette, alcohol, or marijuana use. In some cases, local additions to a model may enhance effectiveness. In a study of seven education and criminal justice projects, additions to the program model were positively related to positive outcomes, whereas modifications of the existing fidelity components were not related to greater program effectiveness (Blakely et al., 1987). It also appeared that the greater the number of modifications present, the greater the likelihood that key components linked to effectiveness were changed.

Studies Examining Exposure

Another serious compromise to implementation fidelity is related to program exposure, (sometimes called dosage), the amount of program content received by participants. Although there are some inconsistent findings related to dosage (CPPRG, 1999; Dane & Schneider, 1998), overall, programs are less effective when study participants do not receive the intended dosage (Allen, Philliber, & Hoggson, 1990).

One example of incompleteness of delivery occurs when youth fail to complete the treatment. For instance, comparisons of youth completing the Family Empowerment Intervention (FEI) program (58%) versus those that did not complete the FEI intervention (42%) revealed that youths completing the program had lower rates of delinquency (self-reported crimes against persons and total delinquency, general theft and index crimes) and drug use (drug sales and frequency of getting very high or drunk on alcohol, frequency of marijuana and hair test marijuana) than youths not completing the FEI (Dembo et al., under review).

In other examples, there is a failure by the implementers to deliver the program in its entirety. In the Midwestern Prevention Project, the differences between high and low implementation schools, as measured by amount of implementation or program exposure, was greater than the differences between the treatment and control schools for all measures of substance use. Additionally, the percentage of change in prevalence rates (i.e., proportion of youth using substances) from baseline to one year for cigarette, alcohol, and marijuana use was lowest in the high implementation

schools and highest in the low implementation and no implementation schools. For example, in the most dramatic example of change, the percentage of students who self-reported smoking within the past month *increased* from 13% to 24% for the control group, *increased* from 13% to 20% in low implementation schools, but *decreased* from 15% to 14% in high implementation schools (Pentz, et al., 1990).

A comprehensive middle school program aimed at increasing social competencies, social bonding, and school success, using program components that had previously demonstrated success in reducing problem behaviors, failed to achieve the expected levels of implementation. The percentage of students receiving the dosage standard set for each component ranged from 0% to 67%, with a mean of 28% across all components. Effects on youth behavior or attitudes were absent (Gottfredson, Gottfredson, & Skroban, 1998).

In a study of the Bullying Prevention Program, those classes that showed larger reductions in bully/victim problems had implemented the three classroom components of the program to a greater extent than those with smaller changes (Olweus & Alsaker, 1991).

In the Resolving Conflict Creatively Program (RCCP), children whose teachers exposed them to a high number of lessons did better than children exposed to low lessons or no lessons. Involvement in a high lessons RCCP classroom slowed down by an entire year the normal developmental growth in hostile attributional biases and aggressive interpersonal negotiation strategies and the decline in competent interpersonal negotiation strategies.

As important as it is to receive the intended number of sessions, implementing staff are often responsible for compromising the dosage because of their own lack of time, commitment, and resources, undermining any effects that the program may have. Although a common theme uttered by implementers is that some exposure to the program elements is better than none, in the case of the RCCP program, children in the low lessons group fared worse than children who received no lessons (Aber, Jones, Brown, Chaudry, & Samples, 1998).

Quality of Program Delivery and Participant Responsiveness

Fewer studies have examined how the quality of program delivery and participant responsiveness impact program outcomes. One study utilized both of these measures of program integrity to create an integrity index, consisting of ratings by program specialists, who taught the program to students, on eight items: (1) program specialist enthusiasm, (2) the degree to which the delivery met the goals of the program, (3) the degree to which program specialist involved all versus a selected few students, (4) classroom control, (5) class enthusiasm, (6) students' responsiveness, (7) students' degree of participation, and (8) the overall smoothness of the lesson. These ratings of program integrity were found to significantly moderate outcomes for three of seven mediating variables (Hansen, Graham, Wolkenstein, & Rohrbach, 1991).

Conclusion

It is clear from these empirical studies that implementation fidelity is important in achieving successful outcomes. In short, these studies show that the closer an intervention adheres to the original design, the greater the degree of behavior change. These findings underscore the need to understand and document the reasons leading to poor implementation and improve the conditions that can facilitate a high quality implementation.

High quality implementation is more likely when core program components are defined in advance and then systematically monitored to ensure compliance (CSAP, 1997; Gresham et al., 1993). The Blueprints Initiative emphasizes, through each Blueprints book (Elliott, 1997), all core program components and includes a chapter on implementation issues to help sites considering the adoption of a program to think through some of the obstacles that they may face.

Federal and state agencies and private foundations should not be content to just provide money to implement a best practice program. Funding should also be provided to organizations that can help agencies and communities to identify empirically-supported programs, assess the needs of the

site, and help them to select an appropriate program, educate all key players as to the core components of the program and the need for quality implementation, and then help the site to implement the program with fidelity by providing training and technical assistance and some standard for assessing fidelity of implementation. Fidelity assessment instruments, designed by the program developers or the funding agency in collaboration with the program developers, should be used to provide feedback to the implementing agency and the funders. In large-scale projects, monitoring can be done by an outside agency, as CSPV has done in the Blueprints project. Many of the Blueprints replication sites and Blueprints designers have appreciated this role being assumed by our organization. TA Providers have not wanted to assume this role, preferring to devote their time and energies to providing needed support through training and technical assistance. Ambivalence among TA Providers regarding fidelity may also arise in the ongoing struggle to balance fidelity with the stated needs or demands of the site for adaptation. This struggle can become especially difficult if it appears that dissatisfaction with the program is emerging. Implementing sites generally do not have ample time and resources to devote to the monitoring process and often don't recognize deviations or understand the importance of fidelity to all program components. Thus, the monitoring role is often more easily assumed by an independent agency, since it has the least amount of conflicting interests and possesses the technical expertise to help community providers plan for implementation, and develop and sustain a program infrastructure that will exist after the TA providers and evaluators are gone.

Our own experience in this role has shown that adherence to a program can be increased by having an outside agency monitor implementation. For example, the Life Skills Training program had an average implementation coverage of 68% in a clinical trial monitored by the program designer (Botvin et al., 1990). The intervention material covered in the replications of Life Skills Training in 70 sites, monitored by Blueprints, ranged from an overall average of 81% to 86% (three cohorts over two years), a substantial increase over that found in the clinical trial conducted under the most favorable of conditions. This suggests that implementation fidelity can be greatly enhanced in real world applications of programs, where conditions are not usually as favorable for achieving a comprehensive implementation.

While some implementing sites may at first resent the attention being paid by the monitoring agency to all the details of the project, over time, most sites learn to appreciate the higher quality implementation they eventually achieve and its impact on outcomes. One of the Blueprints site coordinators stated the following:

There was a demand for attention to "dotting the i's and crossing the t's" which was beyond the norm. Initially, this attention to detail was experienced as overly anal and as a pain in the [delete]. Eventually, however, over the course of a year of implementation, it became crystal clear that fidelity to the model is as important as the model itself. Now I can see that it is this painstaking fidelity that makes [name of program] (and undoubtedly the other Best Practices modalities) a truly effective program with juvenile offenders and their families.

This sentiment was also reported in the CSAP replication initiative of 16 projects replicating 11 model programs. Most of the replicating agencies felt an obligation to implement with high fidelity. One site had complained to CSAP staff about several of the core program components. Yet, six months later they were grateful for the encouragement to stick to the original program design, because the results promised by the program materialized in a robust way (Gray et al., draft).

Because implementation quality is related to program effectiveness, it is important to identify and understand the factors that impede and enhance high quality implementation. Designers, implementers, and sponsors of programs all share responsibility for implementation quality, and they must work together to develop strategies to facilitate and enhance implementation. "The key to understanding how successful research can be translated into successful practice lies in understanding how programs and policies can be implemented so that quality is maintained and the programmatic objectives intended by the program developers are achieved" (Dusenbury, Brannigan, Falco, & Hansen, 2001).

The remainder of this report will document ways of improving the quality of an implementation, beginning with assessing the feasibility of implementing a program at a site and continuing with the challenges of implementing a program and the conditions that lead to successful implementation.