A review of research on fidelity of implementation: implications for drug abuse prevention in school settings

Linda Dusenbury, Rosalind Brannigan, Mathea Falco and William B. Hansen

Abstract
To help inform drug abuse prevention research in school settings about the issues surrounding implementation, we conducted a review of the fidelity of implementation research literature spanning a 25-year period. Fidelity has been measured in five ways: (1) adherence, (2) dose, (3) quality of program delivery, (4) participant responsiveness and (5) program differentiation. Definitions and measures of fidelity were found not to be consistent across studies, and new definitions are proposed. While there has been limited research on fidelity of implementation in the social sciences, research in drug abuse prevention provides evidence that poor implementation is likely to result in a loss of program effectiveness. Studies indicate that most teachers do not cover everything in a curriculum, they are likely to teach less over time and training alone is not sufficient to ensure fidelity of implementation. Key elements of high fidelity include teacher training, program characteristics, teacher characteristics and organizational characteristics. The review concludes with a discussion of the tension between fidelity and reinvention/adaptation, and ways of resolving this tension. Recommendations are made for developing a consistent methodology for measuring and analyzing fidelity of implementation. Further, researchers and providers should collaborate to develop ways of introducing flexibility into prevention programs.

Introduction
The field of drug abuse prevention in school settings has made significant progress during the past 25 years in identifying factors that promote and inhibit the onset of drug use, and in developing interventions for achieving its prevention. Promising research-based approaches have been identified by documents such as Making the Grade (Drug Strategies, 1996, 1999) as well as, more recently, in lists published by federal agencies such as the Center for Substance Abuse Prevention (Center for Substance Abuse Prevention, 2001) and the US Department of Education (US Department of Education, 2001a,b). Federal and state policies have also recently changed in order to actively promote the adoption of these programs (Hansen and Dusenbury, 2001). As a consequence, many research-based programs and approaches are now being disseminated.

With increased dissemination, the field of prevention must now face new challenges. As Berman and McLaughlin [(Berman and McLaughlin, 1976), p. 349] observe, ‘the bridge between a promising idea and its impact on students is implementation’; however, ‘innovations are seldom implemented as planned’. Drug abuse prevention researchers have concluded that the success of empirically validated prevention approaches depends, in part, on fidelity of implementation. Thus, 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 program developers are achieved.

As research-based approaches to drug abuse prevention in school settings move to wide-scale dissemination and the issue of quality of implementation becomes increasingly important, it will be crucial for researchers and practitioners to understand fidelity. In a variety of fields including mental health, community psychology and education, the importance of monitoring implementation is undisputed (Scheirer and Rezmovic, 1983; Brekke and Wolkon, 1988).

To help inform drug abuse prevention research about the issues surrounding implementation, we conducted a review of fidelity of implementation by examining the research literature from the fields of mental health, prevention of psychopathology, personal and social competence promotion, education, and drug abuse treatment and prevention. To identify relevant literature, we conducted computer-based literature searches spanning a 25-year period through ERIC, PsycInfo and the American Psychological Association, as well as through the Library of Congress looking for such terms as fidelity, integrity, quality, implementation and adherence. General searches using these terms yielded more than 9000 sources; specifically, the search of ERIC yielded 8115 sources, PsycInfo 1011, the American Psychological Association 29 and the Library of Congress 266. The primary author and a research assistant read titles of the articles in order to determine the appropriateness of each article for the review.

The purpose of this review is two-fold. In the first part of the review we examine the history of fidelity of implementation and what is known about fidelity of implementation as it relates to drug abuse prevention in school settings, concluding with a discussion of the tension between fidelity of implementation and the need for reinvention or adaptation. Second, we outline recommendations that will, if implemented, advance understanding in the field of drug prevention in school settings about how fidelity may be assessed and improved with the goal of assisting the field to meet the current demands for dissemination.

### Historical overview of fidelity of implementation

Fidelity of implementation is one of the less emphasized components of the diffusion of innovation theory. Diffusion of innovation theory (Rogers, 1995) provides a way of understanding the process by which new ideas are put into practice. It has been used to understand ideas in a number of areas, including education, psychology and public health. In the early days of the theory, most of the emphasis was on program adoption. In the 1960s and 1970s the Research, Development and Diffusion (RD & D) model was commonly used by federal policy makers. This model was inspired by the space program, and emphasized the importance of rigorous evaluation and validation in demonstration projects. A basic assumption of this model was that consumers would value the results of these evaluation studies and base their decisions about whether to adopt an innovation on these evaluation studies. Consumers were viewed as largely passive in this model and it was expected that programs would be implemented by consumers as intended by program developers (Rogers, 1995).

The assumptions of this model were called into question beginning in the mid-to-late 1970s [e.g. (Berman and McLaughlin, 1976; Fullan and Pomfret, 1977); also, cited by (Blakely et al., 1987); (Farrar et al., 1979; Rappaport et al., 1979; Blakely et al., 1987)]. These authors argued that various characteristics of individual organizations had a powerful influence on whether or not programs were adopted and the extent to which they would be implemented with fidelity.

Among the early studies to raise questions about fidelity of implementation was the Rand report on the Implementation of Educational Innovation which analyzed federal programs supporting educational innovation (Berman and McLaughlin, 1976). The study assessed the implementation of nationally disseminated educational innovations and found that teacher-proof programs or ‘pure technologies’ did not exist in practice. A central conclusion of this study was that there was a consistent
lack of fidelity in the implementation of school programs.

The Rand report observed three patterns of implementation in innovative educational programs: (1) *cooptation* or adapting the program without any changes in organizational behavior, (2) *mutual adaptation* in which the program is adapted at the same time there are changes in the organization, and (3) *non-implementation and non-adoption* in which neither happened.

High-quality implementation simply did not exist in practice. The Rand report revealed that programs that were mutually adapted were more effective than coopted programs. Only with mutual adaptation did organizational behavior change.

Critics have noted a number of problems which raise serious questions about the conclusions of the Rand report. For example, programs studied in the Rand report could be described more accurately as general policy changes rather than specific, detailed curricula [Datta, 1981, cited by (Blakely et al., 1987)]. In addition, Blakely et al. noted that the instruments used in the Rand study were predisposed toward adaptation or reinvention rather than fidelity; specifically, the instrument designed to measure implementation assessed ‘the extent to which projects met their own goals, different as they might be for each project. They did not actually assess component specific fidelity’ (Blakely et al., 1987). Finally, the measures relied in many cases on self-reports by users and were very global (Fullan and Promfret, 1977). Nonetheless, the Rand study must be widely credited as the first systematic examination of the issue of fidelity in the dissemination of innovations.

During roughly the same time frame as the Rand study, research by Rogers and his colleagues (Rogers, 1977; Rogers et al., 1995) revealed that consumers or ‘local adopters’ reinvented or changed innovations to meet their own needs and to derive a sense of ownership (Blakely et al., 1987). As a result of research during this period, the classical RD & D model was modified to provide a more active process of dissemination (Blakely et al., 1987). This reconceptualization of the theoretical model resulted in a shift in public policy approaches to dissemination. For example, a variety of new practices became commonplace; these included specialized conferences, having potential consumers visit sites where programs were being implemented, designation of state-level professionals charged with supporting change in particular areas and training by developers. These activities were designed to increase awareness and familiarity with new programs as well as improve fidelity of implementation (Blakely et al., 1987).

In the late 1980s the perspective on fidelity could be described as divided between those who would argue for close adherence to program methods and intent [the pro-fidelity camp, e.g. (Boruch and Gomez, 1977)] versus a more moderate position that allowed for reinvention/adaptation short of the ‘zone of drastic mutation’ [the adaptation camp, e.g. (Hall and Loucks, 1977)]. The more moderate position argued that modifications are necessary at sites to address individual needs and that the more flexibility a consumer has to modify the program to meet those needs, the greater the likelihood that a program will be adopted, implemented, institutionalized and in general have positive results (Berman and McLaughlin, 1976).

### The importance of understanding fidelity

Studying fidelity of implementation is important for a variety of reasons, all of which are related to gaining an understanding of how the quality of implementation can be improved when research-based programs are disseminated. First, in studies in which there is a failure to implement the program as planned—known as a Type III error—there is the potential to conclude erroneously that observed findings can be attributed to the conceptual or methodological underpinnings of a particular intervention (Dobson and Cook, 1980). As Yeaton and Sechrest point out, when we understand that an intervention was not implemented with fidelity, data suggesting that it failed to have an effect become ‘totally uninteresting’ (Yeaton and Sechrest, 1981).
A second important reason for studying fidelity of implementation is that it often helps to explain why innovations succeed and fail. If interventions succeed or fail depending on the dose or quality of intervention, this is crucial information.

Third, an assessment of fidelity of implementation allows researchers to identify what has been changed in a program and how changes impact outcomes, i.e., fidelity can often be observed to affect not only primary behavioral outcomes, such as substance use, but to affect mediating variable outcomes such as changes in attitudes and beliefs as well. Understanding how fidelity moderates such effects can be crucial to guiding refinements in interventions.

Finally, fidelity of implementation reveals important information about the feasibility of an intervention—how likely it is that the intervention can and will be implemented with fidelity. If it is difficult to achieve fidelity of implementation in practice, a program has low feasibility. Programs that are implemented with high levels of fidelity but fail to produce desired effects may need to be redesigned.

**Definitions of terms**

Fidelity of implementation refers to the degree to which teachers and other program providers implement programs as intended by the program developers. While there is agreement generally about what is intended when research refers to fidelity, in fact, fidelity has come to refer to a broad and loosely collected set of specific definitions. The diversity of definitions given to fidelity of implementation include (1) strict adherence to methods or implementation that conforms to theoretical guidelines (particularly when the intervention is adapted to meet the needs of specific circumstances), (2) completeness and dosage of implementation, (3) the quality of program delivery (the way a teacher implements a program), (4) the degree to which participants are engaged, and (5) program differentiation (the degree to which elements which would distinguish one type of program from another are present or absent).

A singular term that defines fidelity has not yet emerged. Each of these specific definitions have value and it is important for research projects to be clear about which specific fidelity issues are being addressed.

**Measuring fidelity**

In the last decade-and-a-half researchers have begun to apply systematic methods to measure critical elements of prevention programs (Weissberg, 1990). Blakely et al. report that multiple methodologies for measuring fidelity have been under development since the mid-1980s (Blakely et al., 1987). Still, measures of fidelity of implementation have been weak (Brekke and Wolkon, 1988), no widely applicable standardized methodology exists for measuring fidelity (Waltz et al., 1993), and valid measures of program implementation and dissemination are needed (Basch, 1984). In part, the challenge of developing measures involves not only defining concepts to be measured, but in developing measures that can be used for assessing fidelity for interventions that differ markedly in their approach. Understandably, measures that have been developed have often been specific to the program or policy being assessed.

In general, fidelity of implementation has been measured in five ways (Dane and Schneider, 1998): (1) adherence to the program, (2) dose (the amount of the program delivered), (3) quality of program delivery, (4) participant responsiveness and (5) program differentiation (whether critical features that distinguish the program are present). Table I summarizes measures used in drug abuse prevention studies. Table I does not provide an exhaustive summary of the research; rather, studies included serve as examples of the kinds of measures that have been used. We did not find measures of program differentiation in the literature we reviewed.

Dane and Schneider strongly recommend that researchers measure all five dimensions of fidelity in order to provide a comprehensive picture of program integrity (Dane and Schneider, 1998). It is not clear that all five types of fidelity must be present for a program to achieve its goals. However,
A review of research on fidelity of implementation

there are strong reasons to investigate this question and, without data, conclusions cannot be reached. As yet, no single published study of drug abuse prevention has included all of them.

Drug abuse prevention researchers have used a variety of means to assess fidelity in the past, including logs which allow teachers to document the activities they covered, and observations and site visits during which objective and subjective information is gathered (Drake et al., 1996). A common source of information about fidelity comes from teacher surveys at the end of programs (Gottfredson et al., 1990). As reviewed below, the specific methods employed may need to be adapted to the type of data being collected and the complexity of information being assessed. We fully expect numerous new approaches to be developed during the coming years.

**Adherence**

Many programs consist of both essential as well as non-essential elements. McGrew et al. argue that the first step in assessing fidelity of implementation should be the identification of critical elements of effective programs (McGrew et al., 1994). In drug prevention programs, central elements are often summarized in statements that define objectives. These may then be used to assess adherence to the planned intervention. For example, Botvin et al. sent observers into selected classroom sessions to observe Life Skills Training being delivered (Botvin et al., 1990b). Observers noted which activities were taught and, for the ones taught, how many achieved the objectives that were stated in the curriculum.

One strategy for assessing adherence may be to have teachers self-report about topics that were covered in any given session or activity. However, a recently published report (Hansen and McNeal, 1999) suggests that many teachers may not have sufficient training in prevention concepts to distinguish between included and excluded elements. In the Hansen and McNeal study, paired observers who had been well trained in understanding the content of drug education had high agreement, whereas observer–teacher agreement was low (Hansen and McNeal, 1999). Thus, if completing theory-based objectives forms the basis of assessing adherence, the primary strategy for measurement may need to rely on observation rather than self-report. Further, other studies provide strong evidence for the validity of observer assessments of fidelity of implementation [e.g. (Harachi et al., 1999)]. For the purposes of this review, we define adherence as the extent to which implementation of particular activities and methods is consistent with the way the program is written.

**Dose**

In many research settings, dose or the completeness of delivery is not a concern, primarily because failure to deliver a program may be rare in situations where programs are controlled by researchers and delivered by paid personnel. However, when a program is implemented by non-research personnel, measuring dose (e.g. numbers of sessions completed, duration or intensity) may provide crucial information about fidelity. In the Midwest Prevention Project field trial of the Project STAR curriculum, Pentz et al. assessed how many of the Project STAR activities teachers in the treatment and control condition delivered (Pentz et al., 1990).

Dose can be measured by teacher logs or checklists. Because there may be perceived pressure to perform, self-reports are expected to be over-estimates of actual dose. Allen et al. collected information about dose by asking participants to report the number of hours of programming received, how schools chose to structure the program (e.g. after school, for credit, etc.) and program provider reports of how much of the program they used (Allen et al., 1990).

For the purposes of this review, dose is defined as the amount of program content received by participants. Most studies that have measured dose have looked at the amount of the program covered [e.g. (Pentz et al., 1990)], often relying on teachers’ self-reports; others might argue that dose can be calculated by extrapolating from the amount of coverage per session for a sample of sessions [e.g. (Botvin et al., 1990b)]. However, we would argue that, ideally, a variety of factors should be included...
<table>
<thead>
<tr>
<th>Study Source</th>
<th>Format</th>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Abbott et al., 1998)</td>
<td>observers minute-by-minute checklist</td>
<td>quality</td>
<td>targeted teaching practices delivered with judged fidelity</td>
</tr>
<tr>
<td>(Basen-Engquist et al., 1994)</td>
<td>teacher self-reports teacher survey forms and logs</td>
<td>adherence</td>
<td>percentage of content areas and methods taught within subject area actually delivered</td>
</tr>
<tr>
<td>(Botvin et al., 1994)</td>
<td>student self-reports student surveys</td>
<td>adherence</td>
<td>total number of subject areas taught</td>
</tr>
<tr>
<td>(Botvin et al., 1990b)</td>
<td>observer form developed for each session</td>
<td>adherence</td>
<td>were certain activities done in class</td>
</tr>
<tr>
<td>(Botvin et al., 1990a)</td>
<td>observer form completed at the end of the intervention</td>
<td>adherence</td>
<td>percentage complete (checked if prescribed material was covered, averaged for each session and each teacher)</td>
</tr>
<tr>
<td>(Botvin et al., 1989)</td>
<td>observers forms developed for each session</td>
<td>adherence</td>
<td>subjective rating of the extent to which the teacher followed the protocol</td>
</tr>
<tr>
<td>(Dent et al., 1998)</td>
<td>teacher self-reports evaluation forms at the end of each session</td>
<td>adherence</td>
<td>percentage complete (percent of points and objectives covered)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>quality</td>
<td>qualitative measures of teacher effectiveness including teacher effectiveness and enthusiasm</td>
</tr>
<tr>
<td></td>
<td>student process measures post-program survey</td>
<td>responsiveness</td>
<td>teacher ratings of amount of effort needed to make points and how accurately program was interpreted by class</td>
</tr>
<tr>
<td>(Hansen, 1996)</td>
<td>observers form used for multiple activities</td>
<td>quality</td>
<td>assessed if activity objectives were completed, the effectiveness of teaching methods, and the overall quality of the activity</td>
</tr>
<tr>
<td></td>
<td>students post-program survey</td>
<td>responsiveness</td>
<td>involvement of students in the activity assessed quality of the teacher, student satisfaction with the program and interaction with the teacher (rating)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>quality</td>
<td>assessed active participation (rating) and whether they shared information with others (rating)</td>
</tr>
</tbody>
</table>
Table I. Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Data Collectors</th>
<th>Data Collection Methods</th>
<th>Data Use</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Hansen et al., 1991)</td>
<td>observers, teachers</td>
<td>form used for multiple sessions</td>
<td>quality</td>
<td>ratings of: teachers’ enthusiasm and extent to which instruction met the goals of the program</td>
</tr>
<tr>
<td>(Harachi et al., 1999)</td>
<td>observers</td>
<td>coding forms to assess teaching strategies and behaviors in 90-s intervals</td>
<td>quality</td>
<td>responsiveness</td>
</tr>
<tr>
<td>(Pentz et al., 1990)</td>
<td>teachers</td>
<td>post-program survey</td>
<td>adherence</td>
<td>assessed whether teachers used strategies that contributed to (versus those that detracted from) high-quality implementation and student involvement</td>
</tr>
<tr>
<td>(Rohrbach et al., 1993)</td>
<td>teachers</td>
<td>post-program survey</td>
<td>dose</td>
<td>dose assessed which sessions were implemented</td>
</tr>
<tr>
<td>(Sobol et al., 1989)</td>
<td>observers</td>
<td>videoed sessions rated using study-specific forms</td>
<td>adherence</td>
<td>measured the extent to which curriculum guidelines were followed</td>
</tr>
<tr>
<td>(S菪h et al., 1998)</td>
<td>observers</td>
<td>forms specific for the study</td>
<td>adherence</td>
<td>assessed global quality including how well the activity did and how well the instructor involved the class in discussion</td>
</tr>
<tr>
<td>(Tappe et al., 1994)</td>
<td>teacher self-reports</td>
<td>telephone interviews and surveys</td>
<td>adherence</td>
<td>assessed coverage of key program concepts and detailed group activities</td>
</tr>
</tbody>
</table>

243
into a calculation of dose: (1) self-reports by providers for all lessons, (2) extrapolations based on observations for a sample of lessons to obtain an objective assessment of the proportion of the curriculum covered and (3) attendance data for each participant.

Quality of delivery
Many school-based prevention programs include interactive techniques that guide students to gaining skills or developing specific attitudes and beliefs. More than simply performing from a script, these methods rely heavily on the program developer to act as a facilitator and coach. Indeed, Tobler and Stratton identified interactivity as a key for successful drug prevention (Tobler and Stratton, 1997). The quality of interaction and the degree to which interactive activities focus attention on desired elements are thus important to measure.

Hansen et al. assessed the quality of delivery using both observation and self-report (Hansen et al., 1991). Observers and teachers used a seven-point scale to describe the how well two theory-based programs were delivered. Raters were required to match a normal distribution in awarding their scores (when averaged over numerous observations) in order to eliminate a positive-response bias. While quality of delivery has been defined in a variety of ways in different studies, for the purposes of this review it is defined as ratings of provider effectiveness which assess the extent to which a provider approaches a theoretical ideal in terms of delivering program content.

Participant responsiveness
Several studies have assessed how participants viewed their participation in an intervention. For example, Hawkins et al. measured whether participants in an intervention were aware of intervention components which should have been obvious (Hawkins et al., 1991); the intervention they were assessing involved training teachers to use cooperative learning strategies. These researchers expected students whose teachers received the training to have more experience with cooperative learning than control students. To test this, students in both experimental and control groups were asked to respond to a question about how often their class divided into small groups or teams which competed with each other.

Hansen asked students who had participated in All Stars and DARE about the degree to which they (1) felt their opinions were respected, (2) participated in class discussions, (3) discussed the program with their parents and (4) would recommend the program to others (Hansen, 1996). Similar systematic assessments of participants participation in, reaction to and recommendations about programs can be reliably obtained through self-report. For the purposes of this review we define responsiveness as ratings of the extent to which participants are engaged by and involved in the activities and content of the program.

Program differentiation
Researchers are recognizing the importance of moving beyond the ‘black box approach’ to prevention evaluations to an approach which attempts to explain the specific ways in which outcomes were achieved (Harachi et al., 1999). Measuring program differentiation can be key to assessing aspects of fidelity that are related to immediate outcomes. We define program differentiation as identifying unique features of different components or programs so that these components or programs can be reliably differentiated from one another.

A complicating feature of many drug abuse prevention programs is that they usually include several different components. For example, social resistance skills training approaches include material on resisting advertising, resisting direct peer pressure, correcting misperceptions about social norms and norm setting, and public commitments not to use substances. Competency enhancement approaches often include social resistance skills training and norm setting as well as training in decision making, anxiety management, goal setting, communication, social skills and assertiveness (Hansen, 1992).

Even where elements have not been found in evaluation studies to be effective when used in
isolation (e.g. information dissemination or self esteem enhancement), it is not uncommon to find these elements used in conjunction with more promising prevention elements (e.g. social resistance skills training or competence enhancement). To the extent elements should be excluded (e.g. scare tactics or segregation of high risk youth), program differentiation measures could be useful in ensuring that these elements are not present.

Component analysis has been rare (Donaldson et al., 1994; Hansen et al., 1988; Hansen and Graham, 1991); however, component analysis is important to determining which elements of prevention programs are essential. In fact, the greatest value of measures of program differentiation may be their contribution to components analyses which could be used to determine the essential elements of effective prevention strategies. For instance, in a study by Hansen et al., two program components, norm setting and resistance skills, were compared (Hansen, et al., 1988; Hansen and Graham, 1991). Measures of program differentiation could have been used in that study to ensure that program providers in one condition (i.e. norm setting or resistance skills) did not implement elements of the other condition and that each program changed only targeted mediators.

Fidelity in drug abuse prevention research

In many studies, fidelity of implementation has been associated with improved student outcomes (Botvin et al., 1990b; Battistich et al., 1996; Abbott et al., 1998; Haynes, 1998; Whitehurst et al., 1999). In the prevention literature in school settings, most researchers view any changes in a program as a potential threat to the integrity of their intervention, and most research has supported the expectation that the more completely a teacher implements a program, the less likely students will use drugs [e.g. (Botvin, et al., 1990b; Resnicow et al., 1993; Rohrbach et al., 1993)]. The inverse also has been shown to be true—when programs are not implemented as intended they are less likely to be effective. In addition, fidelity of implementation has been associated with changes in mediating variables believed to be responsible for outcomes (Hansen et al., 1991).

Projects assessing fidelity

In the social sciences, the literature on fidelity of implementation has been limited. In 1973, Pressman and Wildavsky expressed surprise at the paucity of literature (Pressman and Wildavsky, 1973). Table II provides a summary of the results of reviews of fidelity of implementation in a variety of research fields over the past 25 years. For example, studying the peer-reviewed prevention literature from 1980 to 1994, Dane and Schneider report that only 24% of 162 studies evaluating the effectiveness of primary and secondary prevention of behavioral, social and academic problems assessed fidelity of implementation, and only a third of those considered the impact of fidelity on outcome (Dane and Schneider, 1998).

The proportion of studies in the prevention literature which assessed fidelity was slightly higher than in the treatment literature. For example, a study by Peterson et al. from 1968 to 1980 reported that only 20% of 539 studies in behavioral analysis assessed fidelity of implementation (Peterson et al., 1982). Moncher and Printz reported that 18.1% of clinical treatment studies assessed program integrity or fidelity of implementation (Moncher and Printz, 1991). A review of parent training studies showed that only 6% of studies conducted between 1980 and 1988 reported fidelity of implementation [(Rogers-Weise, 1992); cited by (Dane and Schneider, 1998)]. Fifteen percent of behaviorally based interventions from 1980 to 1990 reported about fidelity of implementation [(Gresham, et al., 1993); cited by (Dane and Schneider, 1998)]. For the purposes of this report it is interesting to note that Dane and Schneider’s review revealed that the research field most likely to consider impact of fidelity of implementation was the substance abuse prevention literature (Dane and Schneider, 1998).

To date, fidelity of implementation has not been extensively studied or reported in the research literature. Researchers often report the measures they took to promote fidelity of implementation, but do not describe methods used to assess it (Caplan et al., 1992; Waltz, et al., 1993). When
<table>
<thead>
<tr>
<th>Citation</th>
<th>Scope</th>
<th>Goals</th>
<th>Definition of fidelity</th>
<th>Measures</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(Berman and McLaughlin, 1976)</em></td>
<td>reviewed 293 educational interventions completed in 1974; conducted 1735 interviews with practitioners</td>
<td>to identify factors influencing outcomes</td>
<td>'The extent to which the project was implemented as originally planned'</td>
<td>mutual adaptation in which both program and participating organization changed; non-implementation in which neither changed; cooptation in which program was changed, but not participating organization; technological learning in which the participants changed but not the program</td>
<td>technological learning did not appear in practice, thus there was an absence of high-fidelity implementation; mutual adaptation, in which local sites adapted programs to meet their needs, produced the most effective result</td>
</tr>
<tr>
<td><em>(Blakely et al., 1987)</em></td>
<td>conducted site visits to examine seven educational and criminal justice programs being nationally disseminated</td>
<td>to refine measures and methods for assessing implementation</td>
<td>fidelity of implementation refers to adherence to model according to the developer; reinvention as (1) addition—new component and as (2) modification of existing component</td>
<td>detailed process measures to identify program components and assess adherence to program model; the component variations were rated as ideal, acceptable or unacceptable</td>
<td>all sites implemented within acceptable range for fidelity of implementation; high fidelity was associated with more positive outcomes; local modifications did not moderate effectiveness; local additions increased effectiveness</td>
</tr>
<tr>
<td><em>(Dane and Schneider, 1998)</em></td>
<td>reviewed 162 primary and early secondary prevention programs of behavioral, social and academic problems, published between 1980 and 1994</td>
<td>to assess the degree to which fidelity of implementation was measured and promoted in evaluation studies</td>
<td>'The degree to which specified procedures are implemented as planned' (p. 23)</td>
<td>adherence, exposure, delivery quality, responsiveness as measured by (1) implementers and (2) trained observers, distinguishing features</td>
<td>39 out of 162 studies systematically assessed fidelity of implementation; 13 of 39 analyzed impact of fidelity on outcomes; lower fidelity ended to be associated with reduced effects</td>
</tr>
<tr>
<td><em>(Emrick et al., 1977)</em> [report cited by (Rogers, 1995)]</td>
<td>national survey of educational interventions through the National Diffusion Network.</td>
<td>to assess the extent to which organizations modify interventions</td>
<td>not specified</td>
<td>not specified</td>
<td>over half (56%) of organizations made some modification; only one in five made major changes</td>
</tr>
<tr>
<td>Citation</td>
<td>Scope</td>
<td>Goals</td>
<td>Definition of fidelity</td>
<td>Measures</td>
<td>Conclusions</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>--------------------------------------------</td>
<td>------------------------------------------------</td>
<td>------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(Gresham, 1993)</td>
<td>review of behaviorally based interventions from 1980 to 1990</td>
<td>to determine the extent to which integrity was assessed in studies</td>
<td>integrity was the extent to which a program is implemented as intended</td>
<td>rated whether study reported that integrity was measured and whether study reported data on integrity</td>
<td>26 (14%) assessed fidelity of implementation</td>
</tr>
<tr>
<td>(Larsen and Agarwala-Rogers, 1977) [report cited by (Rogers, 1995)]</td>
<td>review of 104 mental health interventions</td>
<td>study adoption of innovations</td>
<td>not specified</td>
<td>not specified</td>
<td>reinvention was slightly more likely to occur than not</td>
</tr>
<tr>
<td>(Moncher and Printz, 1991)</td>
<td>review of 359 clinical treatment studies from 1980 to 1988</td>
<td>to determine the extent to which fidelity is monitored</td>
<td>fidelity refers to the extent to which a program (or independent variable) is implemented as planned</td>
<td>assessed whether the study included use of manuals, supervision of providers, as well as checks on adherence</td>
<td>55% ignored program integrity; one out of eight used all three measures</td>
</tr>
<tr>
<td>(Peterson et al., 1982)</td>
<td>review of 539 behavioral analysis studies from 1968 to 1980</td>
<td>to determine how many articles assess the occurrence of the independent variable or treatment</td>
<td>integrity of the independent variable (i.e. treatment)</td>
<td>rated whether study assessed integrity and/or needed to assess fidelity</td>
<td>only 20% of studies assessed fidelity of implementation</td>
</tr>
<tr>
<td>(Rogers-Weise, 1992)</td>
<td>review of 148 published parent training studies from 1975 to 1990</td>
<td>to review parent training interventions</td>
<td>treatment integrity</td>
<td>used Kratochwill’s (1985) specification for measuring integrity</td>
<td>only 6% of studies reported measuring fidelity of implementation</td>
</tr>
</tbody>
</table>
fidelity has been assessed in treatment studies, it is not uncommon for studies to rely on the self-reports of therapists or clients, or if observers are used, to base assessment of fidelity on a single session (Waltz et al., 1993). In addition, even when studies have assessed fidelity of implementation, it is rare that they discuss the validity or reliability of the measures they use to make those assessments. For example, in a review of the community mental health literature (Scheirer and Rezmovic, 1983), 65% of the studies reviewed (n = 74) that attempted to monitor fidelity did not discuss the psychometric properties of the measures they used (Brekke and Wolkon, 1988).

Levels of observed fidelity

Research conducted about the adoption of educational innovations conducted by the National Diffusion Network revealed that over half (56%) of the organizations that adopted innovations ended up modifying interventions (i.e. implemented only certain parts), although only one in five made major changes [(Emrick et al., 1977); cited by (Rogers, 1995)]. In terms of mental health interventions, reinvention was slightly more likely to occur than not (55 reinvented; 49 were unchanged) [(Larsen and Agarwala-Rogers, 1977); cited by (Rogers, 1995)].

In drug abuse prevention research, the few studies that have assessed fidelity of implementation under real world conditions (i.e. with teachers or other non-research staff delivering the program) have revealed that there is a noticeable deficit in the fidelity of program delivery that is achieved. Good evaluations of drug abuse prevention programs have been done in the context of rigorous field trials where there is considerable effort to get teachers to teach programs exactly as intended. However, even under these circumstances there is tremendous variability in how consistently different teachers present program material. For example, Tortu and Botvin report that teachers in their studies implemented between 44 and 83% of the curriculum points and objectives, with an average of 65% (Tortu and Botvin, 1989). Further, a major evaluation of the Life Skills Training program (Botvin et al., 1990b) found that one in four students had teachers who implemented less than 60% of the important points and objectives in the program.

Rohrbach et al. (Rohrbach et al., 1993) conducted an assessment of program dissemination among teachers in schools that participated in the Adolescent Alcohol Prevention Trial (Hansen and Graham, 1991). In this study, teachers from schools that had previously participated in this research project were provided free materials and training, and encouraged to continue to deliver the program. Fewer than four in five teachers who had been trained to implement the program did so. Among teachers who implemented the program in the first year, an average of 75% of the program was delivered. Maintenance was not sustained, however; only one in four teachers taught any of the program in the second year.

Pentz et al. assessed the degree to which teachers in the Midwest Prevention Project adhered to the prescribed set of lessons that constituted Project STAR (Pentz et al., 1990). They found that 68% of teachers deviated at least slightly from their program, although none reported that they had deviated substantially.

An assessment of implementation in a study of Know Your Body (Resnicow et al., 1993) rated over a third of teachers (37%) as low implementers.

A study of the Teenage Health Teaching Modules (Tappe et al., 1995) assessed the long-term implementation of the program by teachers who had been trained. While most schools continued to use the program (94% of 36 original schools), most teachers (84%) reported that they omitted at least one of the modules. Teachers also reported that they were less likely to use role playing, family communication and community involvement—important elements of the curriculum. Reasons for poor fidelity included lack of time, reassignment of teachers who were originally trained and problems with the curriculum.
A review of research on fidelity of implementation

Key elements of high-fidelity implementation

Research has not yet identified a comprehensive list of critical elements which promote high quality fidelity of implementation. However, researchers have suggested likely candidates for critical ingredients. For example, according to the Rand study (Berman and McLaughlin, 1976), key elements to successful implementation include: (1) adaptive planning which is responsive to the needs of participants, (2) training tailored to local sites, (3) a critical mass of participants (including other implementers) to provide support and prevent isolation, and (4) local material development ranging from creating original materials to repackaging materials to make them more appealing to the local audience. More recently, additional factors which impact fidelity of implementation have begun to emerge from the research literature. These include teacher training, program characteristics, teacher characteristics and organizational characteristics.

Teacher training

Prevention researchers view teacher training as an essential element of program integrity [e.g. (Payton et al., 2000)] and see training as essential to promoting successful implementation of prevention curricula (Dusenbury and Falco, 1995). Educational researchers who have studied the implementation process have long recognized that teacher training and staff development are necessary components of any successful implementation involving curricular innovation (McLaughlin and Marsh, 1978; Patterson and Czajkowski, 1979; Basch, 1984; Fullan, 1985; Perry et al., 1997). However, research examining the specific features of training that promote effectiveness has been very limited; there is little known about how training or staff development actually impacts teacher performance or student outcome (Smylie, 1988; Gingiss, Gottlieb and Brink, 1994). The research that does exist has focused primarily on the question of whether some training strategies or components are more effective than others. At best, research on training methods has documented changes in knowledge and, to a lesser degree, attitudes toward prevention approaches.

On the other hand, there has been significant documentation about what teachers prefer to receive when they participate in training. Berman and McLaughlin found that teachers preferred very detailed, concrete instruction in training (Berman and McLaughlin, 1976). Further, having a resource person at the administrative level that had experience with and understanding of the program promoted successful implementation. Outside consultants such as researchers or experts who provided technical assistance were found not to be as useful to teachers.

Documentation of the effects of training on fidelity is limited. A study of Teenage Health Teaching Modules (Parcel et al., 1991) found that teachers who received training were more likely to implement the curriculum with fidelity than teachers who did not receive training. Another study revealed that more extensive training (including follow-up) was associated with higher quality implementation (Smylie, 1988; Perry et al., 1990). Fors and Doster concluded, based on the SHEE evaluation of the Growing Healthy program, that more extensive teacher training was associated with better fidelity and outcomes (Fors and Doster, 1985).

A study comparing live teacher training with video-based training suggested that live teacher training resulted in greater fidelity of implementation (Basen-Enquist et al., 1994) of a smoking prevention program. Specifically, a lower proportion of video-trained teachers implemented the curriculum, although those who did were comparable to the live group in terms of how completely they implemented the program. Video-trained teachers also were less likely to use role plays and brainstorming. However, a separate study by Botvin et al. suggested that live teacher training and video training may result in equivalent levels of implementation (Botvin et al., 1990b). Specifically, Botvin et al. found that levels of program implementation by teachers in a video-based training condition were roughly comparable to level of program implementation by teachers who received
live teacher training (68 versus 67%, respectively) (Botvin et al. 1990b). Training is thus thought of as essential, but specific characteristics that need to be included in training are not yet well understood.

**Program characteristics**

There are a number of potential characteristics that define the structure and operation of a program that have the potential to influence fidelity of implementation. Bauman et al. identified a number of program characteristics which influence fidelity (Bauman et al., 1991). Primary among these is the complexity of the intervention. Specifically, research shows that when interventions consist of many elements that require special skill and that require coordination by many people, they are less likely to be perceived as effective and to be continued by those who use it (Yeaton and Sechrest, 1981). In contrast, programs that are packaged so as to simplify the task of implementation are more likely to be viewed as having a potential to be effective. Essential elements of programs need to be explicitly stated (Fullan and Pomfret, 1977; Gottfredson, 1984). Other program factors include whether or not program instructions are ambiguous, whether the program is sufficiently strong or intense, who sponsors the program and whether it is easy to administer (Yeaton and Sechrest, 1981).

Based on these findings it is clear that detailed instruction manuals on how to implement the program have the potential to enhance fidelity of implementation (Luborsky and DeRubeis, 1984); they allow for greater rigor in delivery and they also allow researchers to assess whether providers adhere to programs. Manuals became a standard in psychotherapy research in the early 1980s and are also a standard in the drug abuse prevention literature. Drug abuse prevention manuals vary considerably in the degree to which they are explicit (Drug Strategies, 1999). In some cases, the methods to be used are explicit, but the goals, objectives and underlying concepts are not. Few manuals explicitly discuss such important issues as how program methods and concepts can be integrated into other instruction topics, and how teachers should respond when there are challenges or when things go wrong.

Because program developers typically have a predefined personal style for writing their manuals, there is little opportunity for the experimental manipulation of manual formats. Nonetheless, teachers may be able to provide clear feedback about a given format. Researchers could take advantage of this feedback and might wish to compare initial impressions about the format with teachers’ impressions after having actually used such materials.

**Teacher characteristics**

A variety of teacher characteristics have been found to predict whether or not a program is adopted or maintained. For example, teacher attitudes toward and support for prevention education have been shown to be related to whether or not a program is adopted or maintained (Gingiss et al., 1994; Parcel, O’Hara-Tompkins et al., 1995). A study by Rohrbach et al. suggested that teachers most likely to continue using prevention programs were newer to the profession, had more training and were more confident in their ability to teach interactive methods, and were more enthusiastic about the prevention program (Rohrbach et al., 1993). Sobol et al. report that teacher characteristics such as confidence and animation during program delivery were associated with adherence and higher integrity, while authoritarianism was associated with lower integrity (Sobol et al., 1989). Another study suggests that teachers’ own smoking status was associated with whether or not teachers were likely to intervene with students who smoked (de Moor et al., 1992).

**Organizational characteristics**

Implementation ultimately depends on the receptivity of the sponsoring organization (Wandersman et al., 1998). A number of organizational characteristics also have been shown to be related to fidelity of implementation, including support by the principal, the teachers’ sense of efficacy as a teacher to educate their students; their ability to communicate; the general school culture; quality of leadership, accommodation and support by administrators; staff morale; whether and to what
extent the organization takes an active approach to problem solving; and the organization’s readiness to adopt new programs (Gottfredson, 1984). Barriers to effective implementation include lack of time, money and other resources. Further, an organization that is overwhelmed or turbulent is likely to have more problems with implementation (Fullan and Pomfret, 1977; Gottfredson, 1984). Studies have examined these characteristics post hoc. It has not yet been determined whether and to what extent organizational characteristics predict fidelity of implementation or, conversely, might be impacted by a successful experience with a new program. Most studies in the field of education have been concerned with preschool or elementary school organizations and little is known about how these findings will generalize to secondary school organizations.

Reinvention versus adherence

As programs are disseminated, the desire to maintain strict adherence and fidelity (primarily held by program developers) is often countered by a desire to adapt, alter and reinvent programs (primarily held by program providers). These conflicting interests have created a tension that, as yet, remains unresolved (Weissberg, 1990). Program developers often possess unique understanding that has guided a program’s development. They understand what are critical elements and what activities are essential to address those elements. They also know what is non-essential and might be easily adapted or omitted without consequence. In many cases, research-based prevention programs have gone through multiple iterations, the early stages of which are often used to correct elements that show little potential for adding to the effectiveness of the program or may even be harmful. Program developers thus fear that adaptation or reinvention may reincorporate elements that should be excluded. Reinvention has the potential to violate theoretical maxims, changing the essential nature of an intervention.

On the other hand, provider organizations fear that strict adherence to some interventions might result in a failure to meet their local needs. Time constraints, community norms, the availability of resources and regulatory restrictions may all play a role in requiring interventions to be adapted. Programs are at risk of not being adopted or continued if they are not designed or cannot be adapted to meet the needs of the provider organization.

In an effort to resolve the tension between the fidelity and mutual adaptation positions, Berman has proposed a contingency model of implementation, arguing that the best strategy (fidelity or adaptation) depends on the nature of the innovation (Berman, 1981). According to this view, the pro-fidelity strategy is most likely to work with highly structured innovations. The adaptation strategy is most effective with less structured innovations. In complex situations, a combination of strategies may be needed.

It is important to note that in the broader social sciences and education there is not agreement about whether strict adherence to a program is a good thing or not. There have been exceptions to the finding that high fidelity is associated with better outcomes [e.g. (Berman and McLaughlin, 1976; McGrew et al., 1994)], and Ridgely and Jerrell raise the question of whether variations from intended implementation are indeed errors, as researchers tend to assume, or whether, instead, they should be viewed as modifications necessary and appropriate to tailor a program to a specific setting (Ridgely and Jerrell, 1996).

The large-scale study of educational innovations conducted by Berman and McLaughlin suggests that mutual adaptation (in which both the program and the organization delivering the program accommodate one another) is the most effective strategy for implementation (Berman and McLaughlin, 1976). In addition, a study by McGrew et al. found that the extent to which teachers modified lessons was associated with greater student efficacy and improved dietary changes (McGrew et al., 1994). Perhaps teachers’ modifications made the lessons more culturally sensitive and appropriate for their particular classroom or perhaps they actually improved the curriculum activities. It could also be
the case that teachers who modified the curriculum were more motivated and creative in general, and thus better teachers. Whatever the explanation, the point remains that modification has been found in some studies to actually improve outcomes.

There is no disputing the concern that adaptation is, at best, a ‘double-edged sword’ (Gottfredson, 1984), bringing with it the possibility that the critical, effective ingredients of a program may be lost when the program is modified to meet the needs of the community. Unfortunately, the research in fidelity of implementation is minimal and flawed (Blakely et al., 1987), and it is difficult currently to have confidence in any of the arguments, for or against fidelity, reinvention or mutual adaptation. One of the major problems today is that research has not yet indicated whether and under what conditions adaptation or reinvention might enhance program outcomes, and under what conditions adaptation or reinvention results in a loss of program effectiveness.

Conclusions and recommendations

Programs implemented as part of research or demonstration projects usually receive considerable support and direction to achieve fidelity of implementation. Outside of research, implementation usually takes place in less than ideal circumstances. Not only is fidelity expected to not be maintained, verifying the degree of fidelity becomes particularly challenging. Nonetheless, as Dane and Schneider point out, understanding fidelity under these conditions becomes crucial for a field of practice to advance (Dane and Schneider, 1998). More information is needed about how to move from research demonstration projects to dissemination on a wide scale (Heller, 1996). In order to advance the quality of research on fidelity of implementation in drug abuse prevention, the following recommendations are made:

(1) The field needs to adopt universally agreed upon definitions of fidelity of implementation. We identified five elements of fidelity that research on programs should adopt that include adherence, dose, quality of program delivery, participant responsiveness and program differentiation (Dane and Schneider, 1998).

(2) Measures of fidelity parallel the five elements of fidelity described above. The development and use of reliable measures and a standard methodology for studying fidelity are needed. We provided examples that typify how each of these elements of fidelity might be measured. However, extensive measurement development is needed, and future studies need to report on each of these in order to understand their relative importance and the inter-relationships that may exist among them. Further, the level of detail currently used to assess fidelity of implementation may not be sufficiently specific; this may be particularly true for community-based and environmental-level interventions. It is also important to consider the type of data used in determining fidelity. Teacher self-reports may be more comprehensive than observer data, but observer data appears to be more valid. At the very least researchers should collect both kinds of data and attempt to validate teacher reports with observational data.

(3) Further research is needed to identify and confirm the factors which influence fidelity of implementation, including: provider characteristics; participant characteristics; the match between providers, participants and the program; and administrative, community and environmental characteristics which influence and promote fidelity of implementation (Morrissey et al., 1997). Research is needed to identify the key elements of high quality implementation. Implementation standards should be developed which specify the minimum criteria for effective implementation including the qualifications of the provider, provider-participant ratios, the frequency and duration of interventions, and the specific content of programs.

(4) An assessment of fidelity of implementation should be strongly encouraged or required in evaluation studies. Funding agencies and
publishers can both help to promote attention to fidelity by adopting minimal acceptable standards for assessing fidelity of implementation as part of their review criteria. Specifically, in order to promote scientific consideration of issues pertaining to implementation and institutionalization, criteria for judging grant proposals should include the researchers’ plans for sustaining the intervention. Similarly, journals should include in their criteria for review assessment of the study’s measures, and analysis of program integrity, fidelity and maintenance (Altman, 1995).

(5) Finally, research in drug abuse prevention provides evidence that poor implementation is likely to result in a loss of program effectiveness. However, experience in disseminating prevention programs suggests that highly detailed protocols that may have been feasible in rigorously controlled research projects may be difficult to maintain in practice. Further, it appears that highly detailed protocols have low success rates in terms of institutionalization. While researchers may be reluctant to relinquish control of prevention programs, it appears that reinvention or adaptation may be necessary to tailor programs to the needs of a particular setting and to promote ownership. Therefore, researchers must begin to develop strategies for increasing the flexibility of programs without compromising their essential core (Altman, 1995). To do this, critical elements of effective programs must be identified and clearly communicated to providers so that providers know what elements cannot be modified. Research should investigate various ways of building flexibility into programs without compromising program integrity. Research is also needed to determine under what conditions programs can be modified. Collaborations between researchers and providers are essential in this process, to identify areas of programs that would benefit from greater flexibility, as well as to determine what types of modifications are acceptable.

Acknowledgements

This project was funded in part by a contract from the National Institute on Drug Abuse, Contract Number 263-MJ-920107, to M. F. (Drug Strategies), Project Director.

References


Center for Substance Abuse Prevention (2001) CSAP Model Programs Website: http://www.samhsa.gov/centers/csap/modelprograms/default.htm


Received on March 6, 2001; accepted on November 7, 2001