

EVALUATION BRIEF

Measuring Implementation Fidelity

October 2009

Introduction

Over the past two decades, researchers and policy makers have increasingly emphasized the need to develop comprehensive evaluations to understand not only the outcomes produced by a program, but also the mechanisms and elements of a program that influence their outcomes. The assessment of implementation fidelity has been highlighted as critical to understanding how programs are implemented in efficacy studies as well as how programs are implemented as they transition from research settings into “real world” settings (Domitrovich & Greenberg, 2000; Durlak & DuPre, 2008; Dusenbury, Branningan, Falco, & Hansen, 2003; Harachi, Abbott, Catalano, Haggerty, & Felming, 1999). As such, procedures for documenting fidelity should be incorporated into program designs and fidelity measurement should be part of evaluation plans.

What is Fidelity?

Fidelity, also referred to as adherence, integrity, and quality of implementation, is the extent to which the delivery of an intervention adheres to the protocol or program model as intended by the developers of the intervention (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000; Mowbray, Holter, Teague, & Bybee, 2003). Researchers have examined and described fidelity across five dimensions: adherence, exposure, quality of delivery, participant responsiveness, and program differentiation (Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury et al., 2003; Fagan, Hanson, Hawkins, & Arthur, 2008):

1. **Adherence:** Program adherence refers to the extent to which program components are delivered as prescribed by the model. Adherence indicators can include program content, methods, and activities. Adherence data are typically reported as the proportion of program components that were delivered compared to the number prescribed. For example, if a provider covered 14 of the 28 content areas of a program, the content adherence score would be 50%.
2. **Exposure:** Program exposure (i.e., dosage) is the amount of program delivered in relation to the amount prescribed by the program model. Exposure can include the number of sessions or contacts, attendance, and the frequency and duration of sessions.
3. **Quality of delivery:** Quality of delivery reflects the manner in which a program is delivered. Aspects of delivery quality can include provider preparedness, use of

relevant examples, enthusiasm, interaction style, respectfulness, confidence, and ability to respond to questions and communicate clearly. The quality of delivery may act as a moderator between an intervention and observed outcomes; for example, if 100% of a program's material is covered but is delivered poorly, positive participant outcomes may not be realized.

4. Participant responsiveness: Participant responsiveness refers to the manner in which participants react to or engage in a program. Aspects of participant responsiveness can include participants' level of interest in the program; perceptions about the relevance and usefulness of a program; and their level of engagement, enthusiasm, and willingness to engage in discussion or activities. Participant responsiveness may play a direct role in outcomes, or may act as a moderator between the intervention and adherence to the program or the quality of service delivery. For example, if participants are not responding well, a provider may omit, modify, or add to the program's content or activities.
5. Program differentiation: Program differentiation is the degree to which the critical components of a program are distinguishable from each other and from other programs. Program differentiation can also refer to the process of identifying the critical components of a program that are essential for producing positive outcomes (i.e., component analysis).¹

Although the majority of fidelity research has focused on adherence and exposure, evaluations should examine all five dimensions of fidelity to provide the most comprehensive picture of program implementation and to help identify those dimensions of fidelity that have the strongest impact on outcomes.

Implementation Fidelity and Program Adaptation

In response to limited funding and budget cuts for child welfare programs, as well as demands for increased accountability, there has been a push for the implementation and evaluation of evidence-based programs. Fidelity to the model underlying an evidence-based program is encouraged in order to preserve the behavior change mechanisms that made the original program effective (Arthur & Blitz, 2000); unfortunately, even when evidence-based programs are adopted, they are often implemented with low levels of fidelity (Elliot & Mihalic, 2004; Ennett et al., 2003; Hallfors & Cho, 2007). There is strong evidence that fidelity levels are significantly related to the amount of positive change achieved by a program. For example, in a review of over 500 studies, Durlak and DuPre (2008) found that mean effect sizes were at least two to three times higher when programs were implemented with high levels of fidelity, especially in terms of adherence and exposure. In addition, studies that incorporate implementation data into outcome analyses often find stronger effect sizes than analyses conducted without these data (Dane & Schneider, 1998; Dusenbury et al., 2003; Harachi et al., 1999; Lillehoj et al., 2004).

Numerous factors affect implementation fidelity, including the characteristics of providers, the organization(s) responsible for implementation, program participants, the community in which implementation occurs, and program support systems (i.e., training and technical assistance) (Carroll et al., 2007; Castro, Barrera, & Martinez, 2004; Durlak & DuPre, 2008;

¹ Several researchers suggest collecting fidelity data from both intervention (treatment) and control groups to allow for measurement of both groups' exposure to the critical components of the intervention (Dusenbury et al., 2003; Mowbray, et al., 2003).

Fagan & Mihalic, 2003; Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005; Hill, Maucione, & Hood, 2007; Mihalic & Irwin, 2003; Pankratz et al., 2006). Although a detailed discussion of these factors is outside the scope of this brief, readers are encouraged to examine the relevant literature in more detail to identify factors that may have the strongest impact on the implementation of their programs.

Despite the documented relationship between fidelity and outcomes, some modifications will inevitably be made to a program and it is unrealistic to expect complete adherence to all components of a program's original model (Backer, 2001; Durlak, 1998). In fact, some adaptation is necessary and possibly beneficial, particularly if the adaptation is in the form of the addition of a new service or activity. If a provider exclusively adheres to prescribed program components in all instances, it may miss opportunities to maximize program effectiveness (Daro & Cohn-Donnelly, 2001). Flexibility in program implementation can also increase program ownership and involvement, thus resulting in longer program life (Backer, 2001; Blakely et al., 1987; Kelly et al., 2000). However, substantial deviations from the original program can become problematic; for example, in a study examining Life Skills Training, Dusenbury, Branningan, Falco, and Hansen (2005) found that while all teachers made adaptations, most were negative changes that detracted from curriculum objectives.

Program adaptation is the deliberate or unintentional modification of a program through (1) deletions or additions (i.e., enhancements) to program components (e.g., to content, materials, activities); (2) modifications to the nature of the components;² or (3) changes in the manner of administration or intensity (i.e., amount or duration) of program components³ (Backer, 2001). Although there should be a balance between fidelity to core components and adaptations to local contexts (Backer, 2001; Durlak & DuPre, 2008), it is not always clear what level of implementation is necessary to achieve positive results. To determine this balance, it is critical that evaluations be designed to collect implementation data, examine the types of modifications made, the reasons for making modifications, and the impact of the modifications on observed outcomes. The assessment of fidelity is particularly important when programs are implemented in real-world settings where program drift is common (Dusenbury et al., 2003).

Why Should Fidelity be Measured?

Researchers have highlighted several important purposes for collecting implementation data (Backer, 2001; Dane & Schneider, 1998; Domitrovich & Greenberg, 2000; Mowbray, et al., 2003; Pankratz et al., 2006).

- *Understanding program implementation:* Implementation data help program staff and evaluators understand how a program was implemented. These data also allow for the examination of deviations from a program model within and across different settings, thereby providing evidence regarding the contexts in which programs are and are not effective.
- *Examining theoretical assumptions:* Implementation data allow evaluators to test theories about the importance of different program components in relation to outcomes.

² For example, a provider may use local child maltreatment statistics instead of national statistics to make the information more relevant to the local context.

³ For example, a provider may conduct activities in a large group rather than in a small group as prescribed by the program model.

- *Interpreting outcome findings:* Implementation data are necessary for understanding why a program succeeds or fails to produce positive outcomes. The data may also elucidate differences in outcomes between sites in multi-site evaluations, ensure internal validity, and help avoid Type III errors.⁴
- *Providing feedback for continuous quality improvement:* Implementation data allow researchers to identify program drift and provide corrective feedback to promote positive outcomes.
- *Providing feedback to program developers about the program:* Implementation data can help identify program components that are consistently implemented poorly. Program developers may use this information to modify components to make them easier to administer or provide additional training to improve implementation quality.

Establishing Fidelity Criteria

Before implementation fidelity can be measured, the critical components of a program's service model must be specified, operationalized, and validated. These "fidelity criteria" can be established in three distinct steps (Mowbray et al., 2006):

Step 1: Identify Fidelity Criteria

The first step in establishing fidelity criteria is to identify the adaptive and core components of a program. The adaptive components are those features of a program that are optional or can be modified to fit the local context without impacting the program's effectiveness. The core components of a program (also referred to as key ingredients, essential elements, or critical components) are those that are theorized to be or are empirically linked to positive outcomes and program effectiveness (Backer, 2001). Core components are those features of a program which must be implemented as faithfully as possible.⁵ The core components of a program should be explicitly stated and operationalized in an intervention manual, curriculum, or other program materials. However, the process of identifying core components can be difficult if a program is not well defined.

Fidelity criteria should reflect all core elements of a program, including content, format, structure, and protocols and procedures. Structural fidelity criteria include features associated with the framework for service delivery, such as staffing levels and characteristics; case load size; supervisory structure; program duration; group size; and the number, frequency, and duration of contacts. There are also process-related fidelity criteria that reflect the manner in which services are provided, such as staff-client interactions, client-client interactions, values, principles, and the emotional climate of the organization (e.g., hostility, chaos) (Mowbray et al., 2003).

Methods for Developing Fidelity Criteria

Several methods can be used to identify fidelity criteria (Emshoff, 2008; Mowbray et al., 2003):

⁴ A Type III error occurs when researchers erroneously conclude that a program is not effective when in fact the failure was due to incomplete or incorrect implementation (e.g., failure to engage participants at an appropriate level).

⁵ Some programs consist of only core components.

1. Draw from a specific program model that has already been explicitly described (and ideally proven efficacious or effective⁶);
2. Conduct component analysis to determine which program components are essential;
3. Examine expert opinions through surveys of experts or literature reviews about effective program components;
4. Conduct qualitative research to examine the opinions of program users and advocates regarding what elements are effective; and
5. Draw from the program's logic model to build upon the theoretical linkages between activities and outcomes.

Step 2: Identify Measurement Tools

The second step in establishing fidelity criteria is to develop and implement methods for measuring (i.e., quantifying) adherence to criteria across all five dimensions of fidelity. The most common methods used to assess fidelity include (1) ratings by experts based on direct observations, videotaped sessions, interviews, project documentation, and client records; and (2) surveys or interviews completed by program staff or participants (Mowbray et al, 2003). Although fidelity is sometimes measured using a single item on a questionnaire (e.g., "how closely did you adhere to the program?"), more accurate and comprehensive results can be achieved by using multiple methods and data sources.⁷

Regardless of which data collection methods are used, program implementation by individual providers should be examined across different program participants and over time because of variations in implementation (Durlak & DuPre, 2008). Although the optimal frequency of fidelity checks depends on the program, they should be conducted frequently enough to prevent program drift but not so frequently that the resulting data lose meaning or the process consumes too many resources (Mowbray et al., 2003).

Provider and Participant Self-Report

Fidelity Assessment Form

One of the most common methods for assessing fidelity involves the completion of an implementation checklist, log, or survey by program service providers. Although most studies currently use paper forms, web-based fidelity monitoring systems are becoming more common (e.g., Hansen, Bishop, & Bryant, 2009; Lee et al., 2008). Some of the advantages of paper or web-based self-report instruments are that they (1) allow evaluators to assess fidelity when they are not present to observe the project themselves; (2) allow for ongoing assessment; (3) place minimal burden on providers; and (4) are less costly than direct observation. Providers can complete fidelity instruments following each program contact or on a daily or weekly basis. Fidelity instruments can track data on specific aspects of implementation, such as:

- Content covered;
- Activities conducted and the time spent conducting the activities;

⁶ Efficacy studies test programs under "optimal," tightly controlled research environments; the results indicate whether a program produces certain outcomes within a specific population and setting. Effectiveness studies test whether an efficacious program produces similar outcomes when implemented under "real world" conditions, with different populations and settings.

⁷ Some evidence-based programs have developed fidelity assessment tools for use when implementing the program (e.g., Fidelity of Implementation Rating System for the Oregon model of Parent Management Training (Knutson, Forgatch, & Rains, 2003) and the Multisystemic Therapy (MST) Therapist Adherence Measure (Henggeler & Borduin, 1992)).

- Methods for delivering intervention;
- Participant attendance; and
- Participant responsiveness to the program.

Fidelity instruments can be created to allow for dichotomous (yes/no) responses, Likert scale responses, or open-ended responses. For programs with discrete sessions or service episodes, a specific session-by-session checklist can track whether each session was conducted (including the date and time of the session) and whether providers covered specific content areas, activities, goals, and objectives associated with a planned session (e.g., Fagan et al., 2008; Harnett et al., 2004; Sanchez et al., 2007). For less structured programs in which providers follow general guidelines rather than adhering to specific content or a specified number of sessions, fidelity instruments may be completed on an ongoing basis or at the end of a program to indicate whether the fidelity criteria were met (Fagan et al., 2008).

Provider Interview Protocols

Interviews with providers are a good way to supplement the fidelity data obtained through self-report or direct observation (Lee et al., 2008). In-person or telephone interviews can be conducted to collect more open-ended data from providers who have a unique perspective about program implementation (e.g., awareness of and reasons for making modifications to a program).

Participant Surveys, Interviews, and Focus Groups

Participant surveys, interviews, or focus groups are good methods for capturing participant responsiveness to a program. Surveys and protocols for interviews or focus groups can be developed to capture information such as how much participants liked the program, the relevance of the content, and participants' perceptions of the service provider.

Self-report measures may be more biased and have lower reliability than observational measures. For example, provider ratings of fidelity tend to be skewed in a positive direction due to social desirability bias (Hansen, Bishop, & Bryant, 2009; Hardeman et al., 2008; Lillehoj et al., 2004). Self-report data may also be limited by the ability of providers or participants to accurately recall information. In addition, providers may not be aware that they are making adaptations to a program (Dusenbury et al., 2005).

Behavioral Observation

Behavioral observations can often provide a more objective assessment of program implementation. For example, observers can assess whether providers delivered the prescribed content, used appropriate delivery methods, engaged participants in the session, or expressed enthusiasm in delivering services. Observational protocols may include scaled rating forms, checklists, or qualitative descriptions of project implementation. Provider self-report instruments can also be used by observers to compare and validate their ratings against those of project service providers. Observers may include members of the evaluation team, agency staff, board members, or members of the community (Fagan et al., 2008). Observers should be trained in the core components of the program model as well as in the procedures for implementing the observation protocol accurately and objectively (e.g., refraining from participating in activities or interacting with participants). As part of the training, observers should practice coding until they meet a certain level of inter-rater reliability.⁸

⁸ A score of .80 is a common standard for an acceptable level of inter-rater reliability.

A comprehensive observation protocol should include a coding manual that defines each program component, outlines the procedures for scoring observations, and specifies the rating scheme (Forgatch, Patterson, & DeGarmo, 2005). Similar to self-report instruments, the rating format for observational tools can include dichotomous responses (e.g., did the provider complete a session or not), open-ended responses (e.g., what adaptations were made to the service delivery methods), or Likert-type scaled items. To maximize objectivity, each point on a rating scale should be anchored to specific behaviors or practices (Mowbray et al., 2003). For example, an item on an observational scale used to assess whether a provider utilized a strength-based approach in providing services could include three scale points: 1 = provider consistently focuses on problems and deficits; 2 = provider points out strengths to family; 3 = provider points out family strengths and builds on these strengths to achieve goals.

Observations can be made either in person or by watching videotapes of program activities. Videotape offers many advantages, such as facilitating observation when distance (or travel costs) precludes direct observation, minimizing intrusiveness, and enabling assessment by multiple observers (Pankratz et al., 2006). The disadvantages of videotaped observations typically involve technological issues, such as depleted batteries or forgetting to press the “record” button during taping (Pankratz et al., 2006).

Some evidence suggests that observational data are more strongly correlated with program outcomes than self-report data (e.g., Dane & Schneider, 1998; Lillehoj et al., 2004), although few studies have directly compared these two strategies (Durlak & DuPre, 2008). However, direct observation is not always feasible given the costs and time required to conduct observational studies. In addition, the act of observation may actually increase bias in certain instances; for example, providers may make more concerted efforts to adhere to a protocol if they know they are being observed, thus creating a skewed impression of typical program performance.

Archival or Administrative Data

Archival and administrative data, such as attendance records, case records, training manuals, and implementation plans, are also useful for assessing fidelity. For example, dosage can be assessed using attendance records to examine the number and duration of participant contacts, while case records can be reviewed to determine if certain activities were conducted (e.g., whether an individualized service plan was created).

Step 3: Assess Reliability and Validity

The third step in establishing fidelity criteria involves assessing the reliability and validity of the criteria themselves. For more information about methods for assessing the reliability and validity of fidelity criteria, see Mowbray et al., 2003.

How are Fidelity Data Used?

Analyzing Fidelity Data

Fidelity data should be compared to pre-established performance targets to determine the degree to which each core component of the program is implemented (e.g., six of eight sessions were conducted). Mowbray et al. (2003) and Durlak and DuPre (2008) recommend against assigning only a single composite score to a program based on the sum of all fidelity criteria because a single score can obscure differences in implementation across providers or programs. For example, composite scores for two service providers could be calculated by

adding up the values for each fidelity criterion. Although the two providers could achieve the same composite score, they could vary dramatically in how they score on certain criteria that have the most significant impact on observed outcomes.

Researchers typically analyze implementation data in two ways (Durlak & DuPre, 2008). The first method involves creating groups of providers based on their level of implementation (e.g., low versus high implementation) and analyzing the data categorically to determine if outcomes differ between the two groups or compared to a control group. A second method that has more statistical power involves assessing implementation using the full range of implementation data (e.g., by using percentages to assess the level of fidelity achieved). Durlak & DuPre (2008) recommend reporting the full range of fidelity values because the designations of “high” and “low” fidelity are often arbitrary and have meaning only in reference to the site in which the data were collected; “high” fidelity at one program site might not be “high” at another site.

Providing Program Feedback

In addition to reporting implementation findings in a final report, evaluators should periodically provide fidelity findings to projects as part of quality improvement efforts. Continuous feedback regarding implementation fidelity allows an agency to take corrective actions as needed (e.g., provide additional training). Several child welfare programs have developed quality improvement procedures. For example, Every Child Succeeds, a multi-site home visitation program in southwestern Ohio and Northern Kentucky, uses a “Red/Green Chart” that indicates whether program performance on certain indicators is meeting or exceeding pre-established targets (colored in green) or is falling short of the targets (colored in red) (see Ammerman et al., 2007). This information guides training and technical assistance for program staff.

Conclusion

Despite some of the challenges described in this brief, the assessment of implementation fidelity is a critical component of a comprehensive evaluation and will greatly enhance your understanding of the strengths of your program and areas in need of improvement. In addition, a strong fidelity assessment will help you interpret your project’s overall findings and identify specific elements of program implementation that have the greatest effect on outcomes. For more information about measuring implementation fidelity, please contact a JBA team member at:

James Bell Associates
1001 19th Street, North, Suite 1500
Arlington, Virginia 22209
(703) 528-3230 or (800) 546-3230
www.jbassoc.com

References

- Ammerman, R.T., Putnam, F.W., Kopke, J.E., Gannon, T.A., Short, J.A., Van Ginkel, J.B., Clark, M.J., Carrozza, M.A., Spector, A.R. (2007). Development and implementation of a quality assurance infrastructure in a multi-site home visitation program in Ohio and Kentucky. *Journal of Prevention & Intervention in the Community*, 34, 89-107.
- Backer, T.E. (2001). *Finding the balance: Program fidelity and adaptation in substance abuse prevention: A state of the art review*. Rockville, MD: Center for Substance Abuse Prevention, Substance Abuse Mental Health Services Administration.
- Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., et al. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology*, 15, 253-268.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40-49.
- Dane, A.V., & Schneider, B.H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18, 23-45.
- Daro, D., & Cohn-Donnelly, A. (2001). Child abuse prevention: Accomplishments and challenges. In J. Myers, L. Berliner, J. Briere, T. Hendrix, C. Jenny, & T. Reid (Eds.), *APSAC Handbook on Child Maltreatment: Second Edition* (pp. 431-448). Newbury Park, CA: Sage Publications.
- Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*, 11, 193-221.
- Durlak, J. A. (1998). Why program implementation is important. *Journal of Prevention and Intervention in the Community*, 17, 5-18.
- Durlak, J.A., & DuPre, E.P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327-350.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W.B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18, 237-256.
- Dusenbury, L., Brannigan, R., Hansen, W.B., Walsh, J., & Falco, M. (2005). Quality of implementation: developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research*, 20, 308- 313.
- Elliot, D., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, 5, 47-53.
- Emshoff, J.G. (2008). Researchers, practitioners, and funders: Using the framework to get us on the same page. *American Journal of Community Psychology*, 41, 393-403.

- Ennett, S.T., Ringwalt, C.L., Thorne, J., Rohrbach, L.A., Vincus, A., Simons-Rudolph, A. & Jones, S. (2003). A comparison of current practice in school-based substance use prevention programs with meta-analysis findings. *Prevention Science, 4*, 1-14
- Fagan, A.A., Hanson, K., Hawkins, J.D., & Arthur, M.W. (2008). Bridging science to practice: Achieving prevention program implementation fidelity in the community youth development study. *American Journal of Community Psychology, 41*, 235-249.
- Fagan, A.A. & Mihalic, S. (2003). Strategies for enhancing the adoption of school-based prevention programs: Lessons learned from the blueprints for violence prevention replications of the life skills training program. *Journal of Community Psychology, 31*, 235-253.
- Fixsen, D.L., Naoom, S.F., Blase, K.A., Friedman, R.M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231). Retrieved November 12, 2006, from http://nirn.fmhi.usf.edu/resources/publications/Monograph/pdf/monograph_full.pdf.
- Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness, and dissemination. *Prevention Science, 6*, 151-175.
- Forgatch, M.S., Patterson, G.R., DeGarmo, D.S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behavior Therapy, 36*, 3-13.
- Hallfors, D., & Cho, H. (2007). Moving behavioral science from efficacy to effectiveness. *International Journal of Behavioral and Consultation Therapy, 3*, 236-250.
- Hansen, W.B., Bishop, D.C., & Bryant, K.S. (2009). Using online components to facilitate program implementation: Impact of technological enhancements to all stars on ease and quality of program delivery. *Prevention Science, 10*, 66-75.
- Harachi, T.W., Abbott, R.D., Catalano, R.F., Haggerty, K.P., & Felming, C. (1999). Opening the black box: Using process evaluation measures to assess implementation and theory building. *American Journal of Community Psychology, 27*, 715-735.
- Henggeler, S.W., & Borduin, C.M. (1992). Multisystemic therapy adherence scales. Unpublished instrument, Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina.
- Hill, L.G., Maucione, K., & Hood, B.K. (2007). A focused approach to assessing program fidelity. *Prevention Science, 8*, 25-34.
- Knutson, N.M., Forgatch, M.S., & Rains, L.A. (2003). Fidelity of Implementation Rating System (FIMP): The Training Manual for PMTO. Eugene: Oregon Social Learning Center.
- Lee, C.S., August, G.J., Realmuto, G.M., Horowitz, J.L., Bloomquist, M.L., & Klimes-Dougan, B. (2008). Fidelity at a distance: Assessing implementation fidelity of the early risers prevention program in a going-to-scale intervention trial. *Prevention Science, 9*, 215-229.

Lillehoj, C.J., Griffin, K.W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior, 31*, 242-257.

Mihalic, S., & Irwin, K. (2003). Blueprints for violence prevention: From research to real-world settings—Factors influencing the successful replication of model programs. *Youth Violence and Juvenile Justice, 1*, 307-329.

Mowbray, C.T., Holter, M.C., Teague, G.B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315-340.

Pankratz, M.M., Jackson-Newsom, J., Giles, S.M., Ringwalt, C.L., Bliss, K., & Bell, M.L. (2006). Implementation fidelity in a teacher-led alcohol use prevention curriculum. *Journal of Drug Education, 36*, 317-333.

Sanchez, V., Steckler, A., Nitirat, P., Hallfors, D., Cho, H., & Brodish, P. (2007). Fidelity of implementation in a treatment effectiveness trial of reconnecting youth. *Health Education Research, 22*, 95-107.